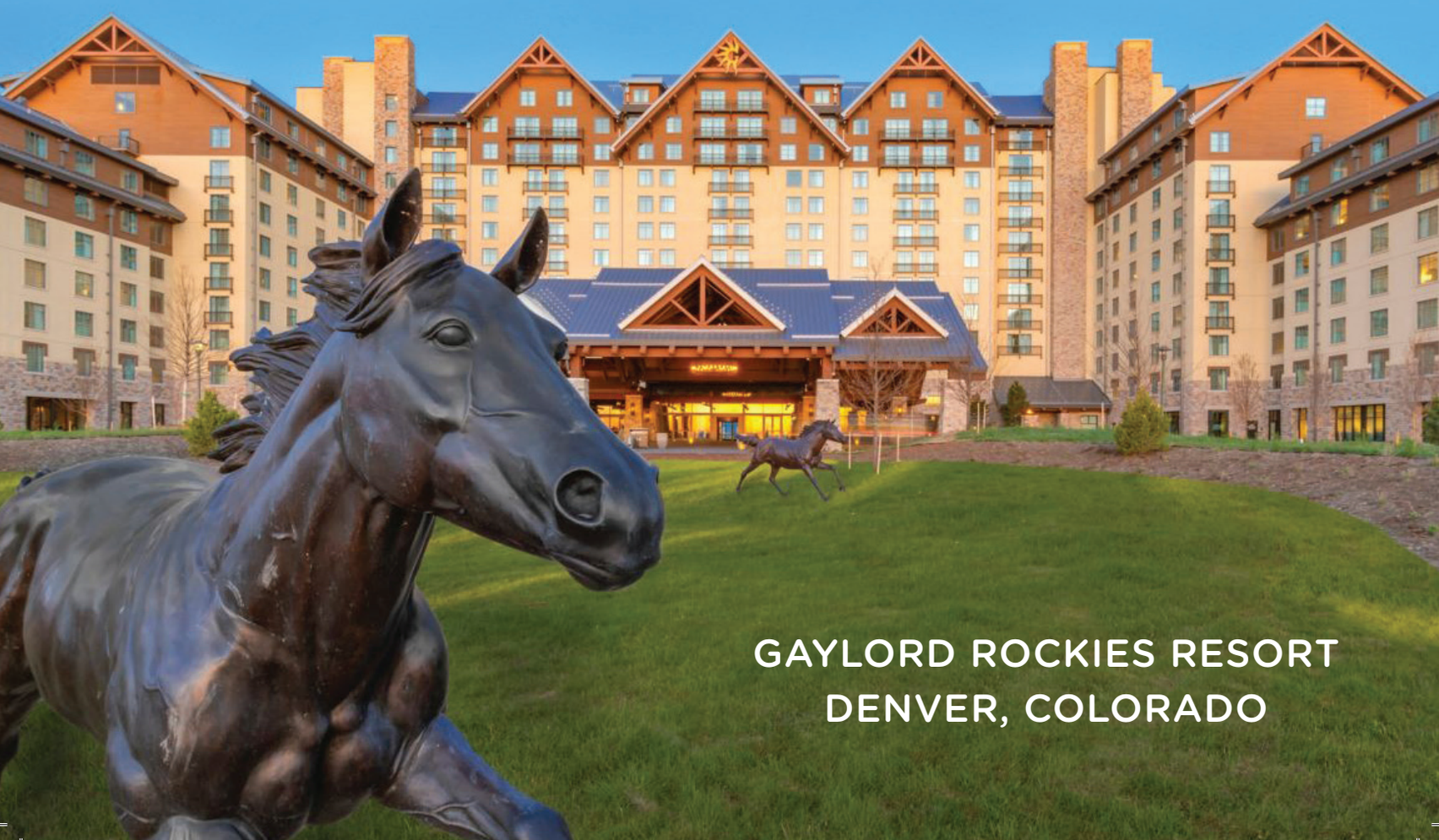
AGBT 2024™

PRECISION HEALTH

SEPTEMBER 4-6, 2024



GAYLORD ROCKIES RESORT
DENVER, COLORADO

WATCHMAKER GENOMICS

OVERCOMING NGS LIBRARY PREP DATA QUALITY BOTTLENECKS RELEVANT TO CLINICAL SEQUENCING APPLICATIONS

JOIN US
THURSDAY
SEPTEMBER 5th
2:15 pm – 2:35 pm
Adams Ballroom B/C/D

Watchmaker is on a quest to maximize data recovery and integrity from challenging samples in sensitive applications, including:

Liquid Biopsy

- Excellent library conversion efficiency delivers increased coverage of rare variants

PCR-free WGS

- Improved coverage of challenging GC-rich regions reduces sequencing costs
- High library conversion efficiency enables inputs as low as 75 ng

FFPE Analysis

- Enzymatic DNA fragmentation reduces sequence artifacts to improve sequencing economy and accuracy
- Rapid total RNA library prep generates libraries in under 5 hours and boosts gene detection with low inputs and FFPE samples

Can't wait? View the data here:

Discover DNA solutions

Explore RNA solutions

Visit us at
Booth 201

watchmakergenomics.com

AGBT 2024™

PRECISION HEALTH

It is our honor to welcome you to the **2024 Advances in Genome Biology and Technology (AGBT) Precision Health meeting.**

Now in its 9th year, AGBT is recognized as a premier gathering for the genomic research community. Since its inception, this meeting has highlighted the most provocative discoveries and major initiatives in genomic medicine from leaders of national healthcare systems, genomics institutes, international biobank initiatives, transformative diagnostic approaches, and hospital-based implementation and reimbursement.

Over the next few days, we will focus on collaboration and dialogue as we aim to pass the torch from pioneering researchers and developers of genome technologies to first movers in healthcare who establish standards of care and cutting-edge applications for genome technology in everyday clinical practice.

AGBT is a not-for profit organization providing three renowned conferences and networking events. Join us for our flagship General Meeting (February 23-26, 2025) and AGBT Agriculture (March 31-April 2, 2025). Learn more about agbt.org.

Organizing Committee

We are incredibly grateful to our meeting Co-Chairs Eric Green, Wendy Chung, Mike Talkowski, and the entire Scientific Organizing Committee for the countless hours they invested in making the AGBT Precision Health 2024 meeting a content-rich experience.

Penelope Bonnen
Baylor College of Medicine

Wendy Chung
Boston Children's Hospital and Harvard Medical School

Catherine Cottrell
Institute for Genomic Medicine, Nationwide Children's Hospital

Geoff Ginsburg
National Institutes of Health, All of Us Research Program

Eric Green
National Human Genome Research Institute

Sue Hill
NHS, England

Stephen Kingsmore
Rady Children's Institute for Genomic Medicine

Howard Mcleod
Utah Tech University

Len Pennacchio
Lawrence Berkeley National Labs, DOE Joint Genome Institute

Heidi Rehm
Broad Institute of MIT and Harvard and Massachusetts General Hospital

Mike Talkowski
Broad Institute of MIT and Harvard and Massachusetts General Hospital

AGBT 2024™
PRECISION HEALTH

We gratefully acknowledge the generous support provided by the following companies

DIAMOND

WATCHMAKER GENOMICS

EMERALD

FABRIC GENOMICS™

SAPPHIRE

Agilent

RUBY

DNAnexus

Exhibiting & Contributing Sponsors

blikka® KURA

Eremid® GENOMIC SERVICES

INOCRAS

ULTIMA GENOMICS

Washington University in St. Louis SCHOOL OF MEDICINE

AlidaBio

General Information

Registration / Hospitality Desk

The AGBT Hospitality Desk will be open in the Cottonwood Foyer during the following hours. Please stop by upon your arrival to pick up your name badge.

- Wednesday, September 4: 11:00am – 7:30pm
- Thursday, September 5: 7:30am – 7:30pm
- Friday, September 6: 7:30am – 5:00pm

Name Badges

Please wear your name badge and lanyard to all AGBT events, including meals. Security will only admit attendees to the conference identified by the AGBT name badge.

Hotel Information

Gaylord Rockies Resort & Convention Center
6700 N Gaylord Rockies Blvd
Aurora, CO 80019
Tel: (720) 452-6900

Transportation

Airport transfers are the responsibility of the individual attendee. Denver International Airport is located seven miles from the Gaylord Rockies Resort. Transfer time is approximately 11 minutes, pending traffic. We recommend using Uber or Lyft as an economical option for airport transfers.

Social Events

- AGBT Welcome Cocktail Hour in Adams Ballroom A: **Wednesday, September 4, 4:00pm – 5:00pm**
- AGBT Welcome Dinner at Grand Lodge Lawn: **Wednesday, September 4, 7:00pm – 10:00pm**
- AGBT Dinner at Front Range Lawn: **Thursday, September 5, 6:45pm – 9:00pm**

Conference Technology

Download our official AGBT Precision Health mobile app to access the agenda, speaker information, maps, abstracts, and notifications. We will notify you through the app if there is a change due to weather.

#AGBTPH
Start Posting!

Agenda

Wednesday, September 4, 2024

11:00am – 7:30pm	**Hospitality Desk** Cottonwood Foyer **Pre-Conference Workshops** Adams Ballroom B/C/D
1:00pm – 2:30pm	**Navigating Clinical Genomic Databases** **Co-Chairs:** Heidi Rehm, co-director of the Program in Medical and Population Genetics, Broad Institute of MIT and Harvard and Massachusetts General Hospital, and Anne O'Donnell-Luria, co-director of the Center for Mendelian Genomics at the Broad Institute of MIT and Harvard **Description:** Participants will learn about the latest advances in clinical genomics databases such as gnomAD, ClinGen, and GenCC and how to use these data for variant discovery, functional prediction, and clinical interpretation. The Co-Chairs will provide live demonstrations of the utility of these databases and address questions. Adams Ballroom B/C/D
2:30pm - 4:00pm	**Leveraging the *All of Us* Researcher Workbench to Advance Precision Health** **Chair:** Anjene (Anji) Musick, Director of the Scientific Data Strategy Branch in the Division of Medical and Scientific Research, National Institutes of Health *All of Us* Research Program **Description:** Participants will learn about the extensive data available on the *All of Us* Researcher Workbench and through a guided tour, complete an example genomic analysis. Adams Ballroom B/C/D
4:00pm - 5:00pm	**Welcome Cocktail Hour – Exhibit Hall** Adams Ballroom A

Session I:	**The Current Landscape of Genomic Medicine (Eric Green, National Human Genome Research Institute and Mike Talkowski, Broad Institute of MIT and Harvard and Massachusetts General Hospital, co-Chairs)** Adams Ballroom B/C/D
5:00pm – 5:10pm	**Welcome (Co-Chairs: Eric Green, National Human Genome Research Institute, and Mike Talkowski, Broad Institute of MIT and Harvard and Massachusetts General Hospital, co-Chairs)**
5:10pm – 5:40pm	**Nancy Cox, Mary Phillips Edmonds Gray Professor and Director, Vanderbilt Genetics Institute and Division of Genetic Medicine, Vanderbilt University Medical Center** "The surprising consequences of heritable lab values for which genetics DO NOT predict the outcomes for which the labs are used"
5:40pm – 6:10pm	**Jordan Smoller, Director, Psychiatric and Neurodevelopmental Genetics Unit and Center for Precision Psychiatry, Massachusetts General Hospital** "Precision and Genomic Medicine for Psychiatry: Unmet Needs and Opportunities"
6:10pm – 6:40pm	**Eimear Kenny, Director, Institute for Genomic Health, Professor|Genetics and Genomic Sciences, Mt. Sinai School of Medicine** "Building programs for genomic medicine in diverse
6:40pm – 7:00pm	**Nita A Limdi, (Abstract Selected Talk), Heersink School of Medicine, University of Alabama** "Population screening for actionable genomic conditions and pharmacogenomics in primary care: The Alabama genomic health initiative"
7:00pm – 10:00pm	**Welcome Reception and Dinner** Grand Lodge Lawn & Patio

Thursday, September 5, 2024 **(Morning)**

7:30am – 7:30pm	**Hospitality Desk** Cottonwood Foyer

7:30am – 9:00am	**Breakfast** Adams Terrace
Session II:	**New Technologies and Novel Methods Empowering Precision Medicine (Wendy Chung, Chief of the Department of Pediatrics, Boston Children's Hospital, and Harvard Medical School, Stephen Kingsmore, President and CEO, Rady Children's Institute for Genomic Medicine, co-Chairs)** Adams Ballroom B/C/D
9:00am – 9:30am	**Kyle Farh, Vice President, Artificial Intelligence, Illumina Inc** "AI for precision medicine and drug target discovery"
9:30am – 10:00am	**Aaron Quinlan, University of Utah Professor and Chair, Department of Human Genetics** "Connections between reduced male fertility and elevated germ cell mutation rates"
10:00am – 10:30am	**Melissa Haendel, The University of North Carolina at Chapel Hill and co-founder of the Monarch Initiative and the National Covid Cohort**
10:30am – 10:50am	**Tian Ge, (Abstract Selected Talk) Massachusetts General Hospital** "Real-time dynamic polygenic prediction for streaming data"
10:50am – 11:15am	**Coffee Break/Exhibit Hall** Adams Ballroom A
Session III:	**The Great Debate: Prenatal and Newborn Sequencing - Are We Ready?** Adams Ballroom B/C/D
11:15am – 12:15am	**Moderator: Heidi Rehm, Broad Institute of MIT and Harvard and Massachusetts General Hospital** **Panelists:** Wendy Chung, Chief of the Department of Pediatrics, Boston Children's Hospital, and Harvard Medical School, Mike Talkowski, Broad Institute of MIT and Harvard and Massachusetts General Hospital, Kelly Gilmore, Genetic Counselor at UNC Health Care

Thursday, September 5, 2024 **(Afternoon)**

12:15pm – 2:00pm	**AGBT Lunch** Adams Terrace
12:30pm – 2:00pm	**Lunch and Learn: The Genome: lessons learned from 7 years of rapid genomes in clinical practice** **Chair, Austin Larson, MD, Children's Hospital Colorado** **Speakers:** **Sara Gracie, Pediatrics-Clinical Genetics and Metabolism, CU Anschutz** "implementation of an inpatient clinical genetics service in the rWGS era" **Kestutis Micke, OB-GYN-Maternal Fetal Medicine, CU Anschutz** "multidisciplinary signout and consultation for prenatal WES/WGS" **Tim Wood, Pediatrics-Clinical Genetics and Metabolism, CU Anschutz** "running a biochemical genetics lab in conversation with molecular genetics and clinical genetics" Adams Ballroom B/C/D
2:15pm – 2:50pm	**Sponsor Talks (Howard McLeod Utah Tech University, Chair)** Adams Ballroom B/C/D
2:15pm – 2:35pm	**Diamond Sponsor - Watchmaker Genomics** **Travis Sanders, PhD, Sr. Research Scientist** "Overcoming NGS library prep data quality bottlenecks relevant to clinical sequencing applications"
2:15pm – 2:50pm	**Emerald Sponsor - Fabric Genomics** **Martin Tristani-Firouzi, M.D., University of Utah** "A Framework to Advance Pediatric Precision Medicine at Scale"
Session IV:	**Transformative Advances in Developmental Disorders (Chairs –Heidi Rehm, Broad Institute of MIT and Harvard, and Massachusetts General Hospital,and Penelope Bonnen, Baylor College of Medicine)** Adams Ballroom B/C/D
2:50pm – 3:20pm	**Elise Robinson, Broad Institute of MIT and Harvard, and Massachusetts General Hospital** "From genes to biology for autism"

3:20pm – 3:50pm	**Jennifer Posey, MD, PhD, FACMG, Assistant Professor, Molecular and Human Genetics, Baylor College of Medicine** "Human and functional phenotyping in developmental disorders: unlocking novel genotype-phenotype connections"
3:50pm – 4:20pm	**Joris Vermeesch, Professor cytogenetics and genome research, scientific director genomics core KU Leuven** "Preimplantation genomic screening to avoid developmental disorders"
4:20pm – 4:40pm	**Sarah Morton, (Abstract Selected Talk) Boston Children's Hospital** "Genomic testing in neonatal hypotonia: an international consortium experience"
4:40pm – 6:40pm	**Poster Session & Exhibit Hall, Wine and Cheese Reception** Adams Ballroom A
6:45pm – 9:00pm	**AGBT Dinner** Front Range Lawn
Friday, September 6, 2024	**(Morning)**
7:30am – 5:00pm	**Hospitality Desk** Cottonwood Foyer
7:30am – 9:00am	**Breakfast** Adams Terrace
Session V:	**Functional and Regulatory Genomics (Chairs Len Pennacchio, Lawrence Berkeley National Labs, and the DOE Joint Genome Institute, and Penelope Bonnen, Baylor College of Medicine)** Adams Ballroom B/C/D
9:00am – 9:30am	**Kristen Brennand, Yale University** "Using Stem Cells to Explore the Genetics Underlying Brain Disease"
9:30am – 10:00am	**Douglas Ruderfer, Vanderbilt University Douglas Ruderfer, Director, Center for Digital Genomic Medicine, Vanderbilt University Medical Center** "Towards precision health: Quantifying physical activity needed to mitigate genetic risk for obesity"

10:00am – 10:20am	**Nicholas Page, (Abstract Selected Talk) University of California, San Francisco** "Large-scale discovery of gene regulatory elements for cis-regulation therapy for autism spectrum disorder and neurodevelopmental delay"
10:20am – 10:40am	**Anne O'Donnell-Luria, (Abstract Selected Talk) Broad Institute of MIT and Harvard** "Noncoding gene discovery for neurodevelopmental disorders"
10:40am – 11:10am	**Coffee Break & Exhibit Hall** Adams Ballroom A
Session VI:	**Fireside Chat – A Patient's Perspective** Adams Ballroom B/C/D
11:10am – 12:00pm	**Moderator: Erica Wright, Assistant Professor & Certified Genetic Counselor, University of Colorado/ Children's Hospital Colorado** "Patient's perspective: a case study"
12:00pm – 12:15pm	**Sapphire Sponsor-Agilent** **Kristi Stephenson** "Gen-Omics and Muti-Omics Solutions from Agilent – Fit for a Diversity of Applications"
12:15pm – 12:27pm	**Ruby Sponsor - DNAnexus** **Asha Collins, SVP & General Manager, Biobanks** "Developing and Using Large- and Small-Scale TREs for Generating Biological Insights" Developing and Using Large- and Small-Scale TREs for Generating Biological Insights. UK Biobank is the largest biomedical resource available to credentialed researchers in the world. The data includes the largest collection of Whole Genome data (500,000 participants) with associated clinical data, the largest proteomics data release in history, as well as data from the largest imaging study conducted to date. The data are exclusively available through the DNAnexus-built, Research Analysis Platform, i.e UKB-RAP. Learn how researchers access and analyze the large-scale UK Biobank dataset and accelerate their analysis from weeks to only hours on UKB-RAP. The session will also cover how the DNAnexus built, UKB-RAP has-

(continued from page 12)

enabled key discoveries in biological insights and precision medicine across such fields as dementia, cardiometabolic diseases, and oncology.

12:27pm – 1:55pm **AGBT Lunch**
Adams Terrace

Friday, September 6, 2024 **(Afternoon)**

Session VII: **Population-Scale Genomic Medicine**
(Chairs - Catherine Cottrell, Institute for Genomic Medicine, Nationwide Children's Hospital and Howard McLeod, Utah Tech University)
Adams Ballroom B/C/D

2:00pm – 2:30pm **Kathleen Barnes, SVP Population & Precision Health, Oxford Nanopore Technologies, University of Colorado**
"The Colorado Center for Personalized Medicine Experience: From Vision to Practice"

2:30pm – 3:00pm **Josh Peterson, Professor of Biomedical Informatics and Medicine, Vanderbilt University Medical Center**
"Integrating genetic and clinical risks: a pathway to better population health"

3:00pm – 3:20pm **Theodore Morley, (Abstract Selected Talk) Vanderbilt University**
"Knowledge informed machine-learning to identify undiagnosed patients with rare diseases in electronic health records"

3:20pm – 3:40pm **Christina Aquilante (Abstract Selected Talk) University of Colorado**
"Preemptive return of clinical pharmacogenomic results to over 50,000 individuals in a population-scale biobank: Phenotype and actionable medication exposure characteristics"

3:40pm – 4:00pm **Meredith Wright, (Abstract Selected Talk) Rady Children's Hospital**
"Prequalification of genomic newborn screening for severe childhood genetic diseases using harmonized population and variant databases"

3:20pm – 4:15pm **Final Comments (Wendy Chung, Chief of the Department of Pediatrics, Boston Children's Hospital, and Harvard Medical School, Eric Green, National Human Genome Research Institute, Mike Talkowski, Broad Institute of MIT and Harvard and Massachusetts General Hospital, Meeting Co-Chairs)**

Agilent supports
AGBT Precision Health

Rachel Larson

Kelly Sheff

Kristi Stephenson

Leave Maybe in the Past –
Choose a Future with Trusted Answers

Visit Booth #101

Agilent

PR7001-2999

25

YEARS OF

AGBT™

AGBT GENERAL MEETING
FEBRUARY 23-26, 2025
JW MARRIOTT
MARCO ISLAND, FL

CALL FOR ABSTRACTS:
SEPTEMBER 24 - OCTOBER 16, 2024

LEARN MORE BY SCANNING THE QR CODE OR VISIT WWW.AGBT.ORG

Abstracts

A framework for large-scale long-read structural variant filtering and integration for All of Us

POSTER 101

Fabio Cunial[1], Ryan Lorig-Roach[1], Adam English[7], Samuel Lee[1], William Harvey[3], Qiuhui Li[9], Yulia Mostovoy[1], Hang Su[1], Steve Huang[1], Karynne Patterson[3], Matt Danzi[4], Josh Smith[3], Chelsea Berngruber[6], Sophie Schwartz[1], Emily Laplante[1], Yuanyuan

1 Broad Institute, Data Sciences Platform, Cambridge, MA
2 Massachusetts General Hospital, Center for Genomic Medicine, Boston, MA
3 University of Washington, Department of Genome Sciences, Seattle, WA
4 University of Miami, Hussman Institute for Human Genomics, Miami, FL
5 National Institutes of Health, All of Us Research Program, Bethesda, MD
6 HudsonAlpha Institute for Biotechnology, Huntsville, AL
7 Baylor College of Medicine, Human Genome Sequencing Center, Houston, TX
8 Vanderbilt University Medical Center, Nashville, TN
9 Johns Hopkins University, Department of Genetic Medicine, Baltimore, MD
10 University of Miami, Miller School of Medicine, Miami, FL
11 Element Biosciences, San Diego, CA

The NIH's All of Us (AoU) Research Program is aggregating electronic health records, physical measurements, and genome sequences for 1M diverse participants. Short-read sequencing (SRS) data has been generated for ~250,000 individuals. However, despite the release of ~1.1B small variants, significant gaps in variant discovery remain. SRS data does not access all genic regions or variation classes equally well, and structural variants (SVs) are particularly under-discovered. To augment this resource, All of Us has initiated several long-read projects (AoU-LR) on ~15,000 participants with multiple technologies and coverage levels (8-15x and 25-30x). A first release of 1,027 participants and ~1M SVs is currently available via the NIH's Researcher Workbench, and a second release of ~5,000 additional samples is scheduled before the end of the year.

Comprehensive ascertainment of SVs requires multiple algorithms with different strengths and weaknesses, presenting significant methodological challenges in filtering and cohort-level integration. Identical challenges arise in other large-scale population genomics and rare disease long-read efforts.

(continue to page 18)

(continued from page 17)

We present a rigorous framework for SV filtering and integration that leverages graph alignment to detect false positives within each sample and integrate calls across samples. We consider SV calls from multiple algorithms as hypotheses in a variation graph and measure how well the reads support such hypotheses. We have trained novel machine learning filters on the samples in the Human Pangenome Reference Consortium (HPRC), using their diploid assemblies as truth and downsampling their PacBio reads to our coverage. Our classifiers achieve an AUC of 0.98, whereas classifiers trained on features from state-of-the-art long-read genotypers achieve an AUC <=0.7. We cast cohort-level SV integration as a combinatorial optimization problem whose

We exemplify the capabilities of the AoU-LR integrated callset, including variants in challenging medically relevant genes and downstream imputation of SVs into large-scale datasets. Since the constraints of our optimization problem can model inheritance in pedigrees, we also demonstrate applications to de novo variation discovery in rare disease trios.

A novel extracellular vesicle–based liquid biopsy method for the early detection of cancer

POSTER 102

Daniel P. Salem[1], Sanchari Banerjee[1], Laura T. Bortolin[1], Ibukunoluwa Po O. Zabroski[1], Delaney M. Byrne[1], Brittany Grimes[1], MacKenzie S. King[1], Lauren T. Cuoco[1], Timothy Santos-Heiman[1], Emily S. Winn-Deen[1], Dawn R. Mattoon[1], Brendan Manning[1], Anthony D. Couvillon1, Sophia Apostolidou2, Aleksandra Gentry-Maharaj2,3, Usha Menon2, Toumy Guettouche1

1 Mercy BioAnalytics, Inc., Waltham, MA
2 MRC Clinical Trials Unit, Institute for Clinical Trials and Methodology, University College London; London, England (United Kingdom)
3 Department of Women's Cancer, Elizabeth Garrett Anderson Institute for Women's Health, University College London, England (United Kingdom)

Detection of cancer early, when it is most treatable, remains a significant challenge due to the lack of diagnostic methods sufficiently sensitive to detect nascent tumors. Early-stage tumors are small relative to their tissue of origin, heterogeneous, and infrequently manifest in clinical symptoms. Detection of their presence is made more difficult by a lack of abundant tumor-specific indicators (i.e., protein biomarkers, circulating tumor DNA, etc.) that would enable detection using a non-invasive diagnostic assay. In addition, many benign conditions manifest similarly, thus discriminating an early-stage cancerous lesion from a benign tumor can present additional challenges and result in unnecessary medical procedures. To overcome these obstacles, we have developed a liquid biopsy assay that interrogates circulating extracellular vesicles (EVs) to detect tumor-specific biomarkers colocalized on the surface of individual EVs. Extracellular vesicles from all cell types, including early-stage tumors, are known to be abundant in blood, are remarkably stable, and serve as a biopsy of their cell of origin. The detection of a colocalized combination of cancer associated biomarkers that provide tumor specificity on the surface of extracellular vesicles enables the discrimination of early- and late-stage cancer from non-malignant conditions. We demonstrate the utility of our extracellular vesicle (EV) based assay for early cancer detection in ovarian cancer (OC). OC ranks among the deadliest gynecological cancers, often diagnosed at advanced stages when treatment options are limited. We evaluated the performance of an EV-based blood test in samples from asymptomatic postmenopausal women in the United Kingdom Collaborative Trial of Ovarian Cancer Screening.

(continue to page 20)

(continued from page 19)

Our assessment included 129 samples from women with high-grade serous ovarian cancer (HGSOC) and 1,310 controls. The OC test showed a sensitivity of 82% (42/51; 95% CI 69%-92%) for HGSOC (all stage), with a sensitivity of 85% (11/13; 95% CI 55%-98%) for Stage I/II. The area under the curve (AUC) was 0.94 for samples drawn 0-12 months prior to clinical diagnosis. The test's specificity was 98% (1,099/1,176; 95% CI 97%-98%. The OC test can sensitively detect HGSOC one year before clinical diagnosis while maintaining high specificity, demonstrating its potential for ovarian cancer screening.

Accurate and robust enumeration of cell type composition from blood and solid tumors with targeted digital cytometry

POSTER 103

Aki Nakao[1], Aaron M. Newman[1,2], Ash A. Alizadeh[1,2], Maximilian Diehn[1,2], Manjula Aliminati[3], Mistuni Ghosh[3], Jayati Ghosh[3], Kristi Stephenson[3], Ashutosh Ashutosh[3]

1 CiberMed, Inc., Saratoga, CA
2 Stanford University, Stanford, CA
3 Agilent Technologies Inc., Santa Clara, CA

Background:
CiberMed's iSortTM software is a widely used digital cytometry solution for profiling cellular composition of blood and tissue samples from bulk RNA-seq. iSort Fractions is a standardized, enhanced version of CIBERSORT/x. iSort Fractions 'blood mode' applies the well-established LM22 gene signature matrix to quantify 22 human hematopoietic cell subsets; iSort Fractions 'tissue mode' quantifies three additional non-immune cell types: fibroblasts, endothelial, and epithelial cells. Agilent SureSelect panels were designed for enrichment of gene signatures from whole blood and tissue. Here we assess the benefits of targeted digital cytometry combining CiberMed's iSort Fractions software with Agilent's SureSelect target enrichment.

Methods:
Whole blood samples from 36 healthy donors and four pairs of matched fresh-frozen (FF) / FFPE samples from non-small cell lung cancer (NSCLC) biopsies were evaluated. Triplicate SureSelect XT HS2 RNA libraries were prepared from total RNA extracted from blood and tumor samples followed by target enrichment with Heme and Tissue panels, respectively. To assess iSort deconvolution accuracy on targeted versus whole-transcriptome bulk RNA-seq data, cell type proportions from complete blood count (CBC, Sysmex) and 6-color TBNK MultiTest (Becton Dickinson, IVD) assays were used as ground truth for blood samples. For tumor samples, deconvolution performance was evaluated by concordance of estimated cell type proportions between matched FF / FFPE samples.

(continue to page 22)

(continued from page 21)

Results:
Cell type fractions in whole blood, quantified by iSort paired with Heme panels, correlated with cell type proportions from complete blood count and TBNK MultiTest assays (r = 0.96) and were reproducible across technical replicates (r ⩾ 0.98). Moreover, this approach reduced the sequencing requirement ~50-fold compared to whole-transcriptome while also improving accuracy. Across 25 cell types from solid tumor samples, cell type fractions determined by iSort paired with the Tissue panel were concordant between matched FF and FFPE samples (rho = 0.93), showing robustness to FFPE artifacts. This approach outperformed deconvolution of whole transcriptome sequencing (rho = 0.78) while requiring ~20-fold less sequencing.

Conclusion:
These targeted panels for digital cytometry are available through Agilent's Community Design program to enable focused, reliable, high-throughput analysis of cell type composition from blood and solid tissue samples.

PR7001-2705

Catalyst by Southern Research: an Alabama-based precision medicine initiative

POSTER 104

Oliver Hampton[1], Kathryn Estes[2], Dionte Rudolph[2], Brooke Bailey[2], Denise Ashimi[2], Khalilah Brown[2]

1 Southern Research, Data Division, Birmingham, AL
2 Southern Research, Medical Affairs Division, Birmingham, AL

Southern Research (SR), an Alabama-based nonprofit biomedical research organization is launching Catalyst, a precision medicine initiative that aims to modernize healthcare through data-driven approaches in the Deep South. For 84-year, SR has translated scientific discoveries into practical applications including the development of 20 novel drugs and collaborations contributing to 50% of all chemotherapies on the market.

Catalyst is a platform that enhances healthcare delivery by providing free genomic sequencing and actionable insights to Alabama patients and their healthcare providers. Through use of actionable insight reports and personalized clinical trial eligibility matching, Catalyst enables tailored treatment based on molecular, clinical, and social characteristics of each patient. Catalyst's actionable insight reports analyze:

- Pharmacogenomics (PGx)
  By analyzing genetic variations that influence drug metabolism and efficacy, PGx ensures treatments are tailored to specific genetic profiles enhancing treatment effectiveness and minimizing adverse reactions.
- Polygenic Risk Scores (PRS)
  By analyzing genetic and clinical factors contributing to disease susceptibility, PRS enables a nuanced understanding of individual health risks.
- Secondary Findings (American College of Medical Genetics and Genomics Genes) By identifying genetic markers associated with various medical conditions, secondary findings analysis allows for early disease detection and intervention.

PGx, PRS, and secondary findings allow patients and providers to develop personalized, proactive intervention strategies that account for drug efficacy, future disease risk, and genetic predispositions. Leveraging genomics, clinical data, and the patient-provider relationship, Catalyst places patients at the forefront of their healthcare decisions. Empowering and enabling individuals with the information and technology needed to actively participate in their health decisions has the potential to mitigate and change poor health outcomes.

(continue to page 24)

(continued from page 23)

Access to precision medicine remains intangible for many, especially minority and rural populations; Catalyst aims to rectify this gap in Alabama. Through statewide, population specific education and engagement efforts for primary care physicians and community members, Catalyst works to build trust between Alabama's marginalized populations, healthcare providers, and genetic researchers to increase diversity within genomic databases and clinical trial enrollments within the populations that need it most.

Characterization of germline cancer predisposition variants in 2500 individuals with pediatric solid tumors using paired exome sequencing

POSTER 105

Ying-Chen Claire Hou[1,2,3], Kareesma Parbhoo[1], Benjamin Kelly[1], Grant Lammi[1], Anthony Miller[1], Vijayakumar Jayaraman[1], Don Corsmeier[1], Douglas Depoorter[1], Prasad Kopparapu[1], Jessica

1 The Steve and Cindy Rasmussen Institute for Genomic Medicine at Nationwide Children's Hospital, Columbus, OH
2 Department of Pathology, The Ohio State University College of Medicine, Columbus, OH
3 Department of Pediatrics, The Ohio State University College of Medicine, Columbus, OH

Identification of germline cancer predisposition variants in pediatric patients allows for cancer surveillance, management, and therapeutic recommendations. Studies have shown that approximately 10% of pediatric cancer patients have an underlying cancer predisposition syndrome; however, the prevalence and spectrum of cancer predisposition variants in children with solid tumors are still largely undefined. Here, we describe the findings of germline cancer predisposition in a cohort of pediatric patients with solid tumors from the National Cancer Institute's Molecular Characterization Initiative. Comprehensive genomic profiling, including paired exome sequencing of a comparator germline sample and disease-involved tissue, was performed across a cohort of 2561 individuals with pediatric solid tumors. The capture panel is comprised of targets for 19,433 genes within the human genome (39 Mb of genomic space) in concert with a genomic "backbone" of hybrid capture probes, enabling the detection of single nucleotide variants (SNV), copy number variants (CNV), and copy-neutral loss of heterozygosity. The cancer distribution in this cohort included central nervous system tumors (71.5%), soft tissue sarcoma (19.0%), rare tumors (8.6%), and neuroblastoma (0.8%). Among this cohort, 341 individuals harbored medically meaningful germline cancer predisposition variation. This included SNV and insertion-deletion (indel) variants in 12% (n=319), and CNV in 1% (n=27). The most commonly involved genes were CHEK2 (n=37 individuals), TP53 (34), DICER1 (23), NF1 (19), PMS2 (18), BRCA2 (16), and ELP1(16). Out of 319 individuals with SNV/indel germline variation, 24 had more than one variant in genes associated with autosomal dominant cancer predisposition.

(continue to page 26)

(continued from page 25)

Somatic second-hit variants predicted to result in loss-of-function of a tumor suppressor gene were observed in 51 individuals in the disease-involved tissue, consistent with Knudson's two-hit hypothesis. Of particular interest for potential new cancer predisposition association, we identified two individuals with pineoblastoma harboring DROSHA germline truncating variants with a somatic second-hit variant in the tumor. Germline alteration in DROSHA has not been associated with a cancer predisposition syndrome; however, biallelic somatic alterations of DROSHA have been described in Wilms tumor and pineoblastoma, thus extending on known disease-gene relationships. In summary, 13% of individuals in this cohort had at least one germline cancer predisposition variant highlighting the significance of paired exome sequencing.

Clinical whole genome sequencing: strategies to develop a comprehensive assay

POSTER 106

Areesha Salman[1], Andrew Bernier[1], Andrea Oza[1], Katherine Lafferty[1], Marina DiStefano[1], Diana Toledo[1], Edyta Malolepsza[1], Samuel DeLuca[1], Katie Larkin[1], Teni Dowdell[1], Justin Abreu[1], Sean E. Hofherr[1], Heidi Rehm[1,2], Niall Lennon[1]

1 The Broad Institute of MIT and Harvard, Cambridge, MA
2 Massachusetts General Hospital, Boston, MA

Broad Clinical Labs (BCL) has sequenced nearly 300k clinical-grade whole genomes (cWGS), and one year ago, leveraged this expertise to launch an expanded service that included patient-specific interpretation and reporting. To-date, over 220 interpretive tests have been completed for over 175 patients, ranging from actionable genomic screening panels for healthy adults to diagnostic sequencing for critically ill newborns.

Since its launch, the clinical WGS test suite has undergone rapid improvement in the laboratory method and the bioinformatics pipeline. Here, we describe the current state of the product and the process of continuously improving a clinical assay in a CLIA-certified, CAP-accredited lab.

One significant example of laboratory method improvement is the change of the underlying sequencing platform from NovaSeq 6000 and S4 Flowcell to the NovaSeq X Plus and 25B Flowcell. This change enabled lower costs and increased throughput, but came with significant changes to flow cell and sequencing chemistry that required optimization and validation. This work describes that process in detail. In brief, a comprehensive validation cohort was gathered including clinical and reference samples. Results of the validation indicated high (>99%) analytical, inter-run, and intra-run repeatability, and high analytical sensitivity. The observed analytical PPV for SNPs and InDels was ⩾ 99.8% and ⩾ 99.5%, respectively, when comparing 5 HG001 replicates to NIST truth data. Among 39 patient samples with known variants of clinical significance, all variants were identified in the NovaSeq X output (18/18 SNVs, 5/5 CNVs, 19/19 InDels, and 4/4 SMN1 copy number variants). Another upgrade has been to increase the types of variants that can be identified from WGS. This includes the validation of the Illumina DragenTM implementation of ExpansionHunter for tandem repeat expansions; a caller for mitochondrial DNA variants; and targeted callers for variants in the PMS2 and NEB genes, which have technical challenges due to high homology. This work is ongoing and each additional variant type requires a rigorous process of collaborative design with the Dragen team, clinical sample procurement, testing, and validation, which will be described herein.

CRISPR/Cas9 genome editing of human B-lymphocytes to produce custom quality-assurance materials for molecular assays used in newborn screening

POSTER 107

David W. Cobb[1], Grace Kim[1], Francis Lee[1], Rachel Lee[1], Carla Cuthbert[1], Rosemary Hage[1], Suzanne Cordovado[1]

1 Centers for Disease Control and Prevention, Newborn Screening and Molecular Biology Branch, Atlanta, GA

Nearly every baby born in the United States undergoes newborn screening (NBS), using dried blood spots (DBS) collected within hours of birth. Extracts from these DBS are used to screen newborns for numerous disorders, ensuring timely and life-saving intervention—so test accuracy is paramount. To support NBS programs in ensuring the accuracy of their molecular tests, CDC produces quality assurance (QA) materials in the DBS matrix. Most lab-created DBS (LcDBS) for molecular testing are prepared using human cells collected from patients or family members with specific genomic variants. However, finding patients with extremely rare variants can be difficult. Additionally, some biomarkers—such as the T-cell Receptor Excision Circle (TREC) used to screen for Severe Combined Immunodeficiency (SCID)—are only present in sufficient quantities in newborn blood. Therefore, production of QA materials containing rare but critical variants or biomarkers is challenging. To begin addressing these difficulties, CDC developed a CRISPR/Cas9 genome editing workflow to create custom cell lines containing the TREC signal joint sequence inserted into the genome of immortalized B-lymphocytes. Post-editing selection methods were then used to establish clonal cell lines harboring the TREC sequence. DNA extracted from these engineered cells performs comparably to DNA extracted from TREC-containing umbilical cord blood cells, thus providing CDC with a sustainable source of cells that test positive for TRECs. Currently, this approach is being used to create an SMN1_E7del+/+ cell line to support NBS for SCID and Spinal Muscular Atrophy (SMA), which is performed using a multiplex PCR assay that detects both the TREC and SMN1 E7del biomarkers. Additionally, this CRISPR/Cas9 workflow is being used to address QA needs for other disorders, by incorporating into cells very rare variants that have been difficult to find in patient/family populations. Establishment of this workflow will allow CDC to develop custom QA materials in various newborn disorders that are inclusive of rare variants, as well as variants associated with underserved populations. This approach to LcDBS production will enable a more dynamic and equitable response to the needs of the newborn screening community.

Deep learning–based integration of tumor omic and functional characterization for precision treatment selection in advanced and metastatic breast tumors

POSTER 108

Casey Sederman[1], Tony DiSera[1], Sandra Scherer[2,3], Chieh-Hsiang Yang[2,3], Emilio Cortes-Sanchez[2,3], Xiaomeng Huang[1], Yi Qiao[1], Bryan Welm[2,3], Alana Welm[2,3], Gabor Marth[1]

1 University of Utah, Department of Human Genetics, Salt Lake City, UT
2 University of Utah, Huntsman Cancer Institute, Salt Lake City, UT
3 University of Utah, Department of Oncological Sciences, Salt Lake City, UT

Despite several successful applications of genome-guided cancer therapy, a lack of rational guidance for treatment selection remains a critical barrier to precision oncology practice. Most tumors lack actionable molecular alterations, limiting the scope of biomarker-informed treatment selection. Further, response to biomarker-informed treatment is highly variable, even among patients harboring theoretically targetable alterations. Collectively, these considerations inform the need to expand guidance for treatment selection to tumors lacking actionable molecular alterations and to drugs lacking validated predictive biomarkers. To this end, we report a hybrid computational/experimental approach to treatment selection, leveraging a clinically oriented functional precision oncology pipeline in concert with a deep learning-based cancer drug response prediction framework, ScreenDL, engineered to capitalize on the patient-specific data generated by this clinical workflow. ScreenDL leverages a phased training schema, incorporating general purpose pretraining in cell lines and follow-up fine-tuning with a newly generated pharmaco-omic resource spanning >100 advanced/metastatic breast cancer (BC) patient-derived xenograft (PDX)-derived organoid (PDxO) models. As patients enter our precision oncology pipeline, omic characterization of the patient's tumor and functional drug screening in patient-derived organoids (PDOs) enables patient-specific fine-tuning of ScreenDL with our ScreenAhead module, generating a personalized response prediction model optimized for the individual patient. ScreenDL outperforms existing DL models in never-before-seen tumor samples, achieving a median Pearson correlation between observed and predicted response across drugs of 0.49 in cell lines and 0.26 in our PDxO cohort. In both contexts, ScreenAhead improves performance, achieving median correlations per drug of 0.68 and 0.54. Further, ScreenDL consistently outperforms the current clinical practice of single- and multi-gene biomarkers for response prediction. Mirroring application in patients and matched PDOs, we applied ScreenDL to select treatments for 7 of the original BC PDX models with matched in vivo screening data. With ScreenAhead, ScreenDL achieves a response rate of 77% amongst selected therapies compared to 44% for the next best tested DL model, demonstrating the vast potential of our computational/experimental strategy for precision treatment selection in the clinic.

Defining kidney cell-type–specific responses in Alport Syndrome through spatial transcriptomic and multiomic integrative analysis

POSTER 109

Haojia Wu[1], Benjamin D. Humphreys[1,2]

1 Division of Nephrology, Department of Medicine, Washington University in St. Louis, St. Louis, MO
2 Department of Developmental Biology, Washington University in St. Louis, St. Louis, MO

Alport Syndrome (AS) is a genetic disorder that affects the kidneys, leading to glomerulonephritis, tubular injury, kidney fibrosis, and end-stage renal failure. Comprehensive profiling of gene expression at single-cell resolution and with spatial context is essential to understand the disease but is currently lacking. Here, we combine the cell-type specific bulk RNA-seq, snRNA-seq and spatial transcriptomics to study the cellular responses in the Alport kidneys.

Col4A3-/- mice were used as a model for autosomal-recessive AS. Kidneys were collected from wild type mice and from Col4A3-/- mice at three time points representing mild, moderate, and severe nephrotic syndrome: Day 42, Day 56, and Day 70. Cell type-specific bulk RNA-seq, snRNA-seq, and Xenium were performed on the same kidneys at each time point. An integrative analysis of the three modalities was conducted to map the cell type-specific responses and the spatial gene expression across the disease time course.

snRNA-seq revealed that that all kidney cell types respond to AS with significant gene expression changes in podocytes, tubular cells, and fibroblasts. Differential gene analysis identified disease-related genes common to all cell types and those unique to specific cell types. Notably, pathway and metabolic activity analyses both indicate defective tryptophan metabolism in the proximal tubular cells (PT) of Alport kidneys. Cell-type-specific profiling of PT confirmed that most genes encoding enzymes for the tryptophan metabolic pathways were downregulated in Alport Syndrome. This defective tryptophan pathway was positively associated with proteinuria, kidney fibrosis, and eGFR decline in both mice and humans. Immunofluorescent staining showed downregulation of the tryptophan proteins such as HAAO and KMO in the fibrotic kidneys of both mouse and human. Targeted in situ spatial profiling by Xenium further supported this defective pathway in PT cells of Alport kidneys. In vitro knockdown of the HAAO gene triggered cell cycle arrest and apoptosis in human primary PT. In addition, protein overload mimicking proteinuria induced downregulation of tryptophan genes, which could be rescued by supplementing NAD+, the end product of the tryptophan pathway.

In conclusion, our multi-omic analysis uncovered a critical disease pathway in the injured PT of Alport kidneys, highlighting potential targets for drug intervention.

Dried blood spot material development and assay optimization to support newborn screening for congenital cytomegalovirus

POSTER 110

Rosemary Hage[1], Jihyun (Grace) Kim[1], David Cobb[1], Francis Lee[1], Suzanne Cordovado[1], Rachel Lee[1], Carla Cuthbert[1]

1 Centers for Disease Control and Prevention, Newborn Screening and Molecular Biology Branch, Atlanta, GA

Cytomegalovirus (CMV) is the most common agent of intrauterine infection in newborns and a leading cause of non-genetic hearing loss in children. Although urine is currently the gold standard for detecting congenital CMV (cCMV) infection, dried blood spots (DBS) have been used as an alternate matrix for detection of cCMV in newborn screening (NBS) since this sample is already collected from all newborns in the U.S. Currently, three NBS programs in North America are screening for cCMV and several programs are in various stages of planning or test implementation. CDC developed CMV-positive DBS materials by spiking CMV-negative human blood with CMV genomic DNA. Results from a lab-developed real-time qPCR assay that amplified CMV genomic targets, UL122 and UL83, and a commercial assay targeting UL122 revealed that these three targets were consistently detected by the CDC laboratory in DBS with viral DNA concentrations simulating low viral load patients. These developmental DBS were shared with three NBS programs to test in their laboratories using their methods. Analysis of the resulting data from these three NBS programs found that all CMV genome targets used by these laboratories were also reliably identified. Further, CDC evaluated several different DNA extraction and PCR assay conditions to optimize a lab-developed real-time qPCR assay for use with DBS that could be adopted by NBS programs. This study lays the groundwork for CDC to support NBS laboratories that screen for cCMV with quality assurance materials and technical support.

Efficient and accurate mixed model association tool for single-cell eQTL analysis

POSTER 111

Wei Zhou[1,2,3,4,5*], Anna S.E. Cuomo[6,7,8*], Angli Xue[6,9], Masahiro Kanai[2,3,10,11,12,13], Grant Chau[2,3], Chirag Krishna[10,11,12], Ramnik J. Xavier[10,11,12], Daniel G. MacArthur[7,8], Joseph E. Powell[6,14], Mark J. Daly[2,3,4,13], Benjamin M. Neale[2,3,5,13]

1 Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA
2 Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA
3 Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA
4 Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland
5 Novo Nordisk Foundation Center for Genomic Mechanisms of Disease, Broad Institute of MIT and Harvard, Cambridge, MA
6 Translational Genomics Program, Garvan Institute of Medical Research, Sydney, New South Wales, Australia
7 Centre for Population Genomics, Garvan Institute of Medical Research and University of New South Wales (UNSW) Sydney, Sydney, New South Wales, Australia
8 Centre for Population Genomics, Murdoch Children's Research Institute, Melbourne, Victoria, Australia
9 School of Biomedical Sciences, University of New South Wales (UNSW), Sydney, NSW, Australia
10 Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, MA

Understanding the genetic basis of gene expression is key to understand the molecular underpinnings of human traits and diseases. eQTL mapping explores this relationship but can vary by cell type and state, necessitating single-cell RNA sequencing (scRNA-seq) and single-cell eQTL (sc-eQTL) analysis. Current sc-eQTL methods either use pseudobulk approaches or do not scale well to large datasets. Here, we propose SAIGE-QTL, a robust and scalable tool that can directly map eQTLs using single-cell profiles without needing aggregation at the pseudobulk level. Additionally, SAIGE-QTL allows for testing the effects of less frequent/rare genetic variations through set-based tests, which is traditionally excluded from eQTL mapping studies.
We demonstrated that SAIGE-QTL is well calibrated for both the single-variant test for common variants and the set-based tests for rare and less frequent variants through simulation studies. We applied it to the OneK1K cohort, which consists of scRNA-seq from over 1.2 million immune cells of 982 unique individuals, categorized into 14 distinct (sub)cell types and matched genotypes.

(continue to page 33)

(continued from page 32)

17,218 eGenes across all cell types (FDR < 5%) were identified by SAIGE-QTL, 5,894 unique genes of which 2,447 (42%) were specific to one cell type only.
We have shown that eQTL mapping at the cellular level using SAIGE-QTL identified 48.4% more eGenes than the previous analyses using the pseudobulk approach. Using a Mendelian randomization approach, we identified more unique loci contributing to autoimmune diseases through gene expression change in one or more cell types. For about a third of the number of eGenes identified through common variant tests, we identified rare (MAF ⩽ 5%) variant eGenes through set-based tests, of which 21% were independent of common eQTL signals from the same genes. In addition, we identified 413 trans- eQTLs for 390 genes in three representative cell types (Naive CD4+T cells, innate B, and plasma cells).
In summary, we developed a novel computational method that addresses the unsolved analytical challenges for systematic genetic analysis on single-cell transcriptomics, offering the potential to interpret complex trait associations through genome-wide colocalization events.

Elucidating the genomic underpinnings of idiopathic scoliosis with whole genome data

POSTER 112

Anxhela G. Gustafson[1], Awtum M. Brashear-Lewis[1], Rinu Kooliyottil[1], Andrew Quitadamo[1], Kamran Shazand[1]

1 The Shriners Children's Genomics Institute, Tampa FL

Idiopathic scoliosis (IS) is defined by the presence of an abnormal structural spinal curvature of greater than or equal to 10º in the coronal plane in the absence of an underlying vertebral malformation. Approximately, 1% of the entire population and 3% of the school-age population is affected with IS. Most affected individuals have mild clinical features with 10% of cases requiring bracing for curves between 25-40º and surgery for curves greater than 50º. Approximately, 25% of patients with IS will fail bracing. Currently, there are no known biomarkers to help determine which patients with spinal curves will progress to bracing and surgery, and which patients will fail bracing treatment. IS is thought to have genetically heterogeneous causes. At the present time, it is not possible for clinicians to utilize genetic knowledge to guide clinical care and treatment. Ultimately, our goal is to use our findings to improve diagnosis and guide personalized treatment approaches for our patients and the community at large.

A whole genome sequencing (WGS) effort is ongoing at Shriners Hospital Centers, which has recruited hundreds of patients with IS (and their parents when available as a trio design). In addition to the genomic data, we inferred phenotypic data with de-identified medical data and consulting clinicians. Using this dataset, we have taken multiple approaches to studying IS including cohort and family-level analysis using Single Nucleotide Variants (SNVs), Copy Number Variants (CNVs), and Structural Variants (SVs). We are currently pursuing a transcriptomics approach to investigate the impact of candidate genomic variants on gene expression and the isoforms involved in target tissues. For functional validation, we aim to test our top candidates in a zebrafish scoliosis model.

Enabling long-read next-generation sequencing for metagenomics applications

POSTER 201

Maurice Chan[1], Yizhen Lin[1], Deng Fong Chao[1], Jiahao Chang[1], Tian-Hsiang Ma[1], Hannah Lucio[1], Hau Hung[1], Mengchu Wu[1]

1 Micronbrane Medical, Hsinchu, Taiwan

Background The advent of low-cost, long-read sequencing by Oxford Nanopore Technologies (ONT) offered the possibility for individual samples to be run on-demand, with theoretical outputs up to 2.6Gb per sample. Herein, we demonstrate the adoption of ONT sequencing for both WGS-based and 16S+ITS-based metagenomics applications. Further, one of the challenges with WGS-based mNGS is the overwhelming amounts of host DNA compared to microbial DNA, reducing the sensitivity to detect microbial sequences. ONT's Rapid Barcoding Kit or the Rapid PCR Barcoding Kit typically necessitates 200ng or 5ng of genomic DNA input, respectively. However, host-depleted DNA often falls below the detection limit of the Qubit assay (<0.1ng/ul). Herein, we show that host-reduced samples were amenable to ONT sequencing using the Rapid PCR workflow.

Methods To set-up the WGS-based and 16S+ITS-based workflows, 5ng of ZymoBIOMICS Microbiome Standard D6331 was used as DNA input. A human blood sample (2.5ml) spiked with 106 SICP (ZymoBIOMICS Spike-in Control D6320) was subjected to host-depletion and gDNA extraction using the Devin Microbial DNA Enrichment Kit (Micronbrane). Extracted gDNA was subjected to library preparation using the Rapid PCR Barcoding kit. Successfully constructed libraries were sequenced individually using Flongle flow-cells.

Results WGS-based and 16S+ITS based workflows resulted in successfully sequenced libraries with comparable analytical performance using D6331. As expected, DNA extracted from the host-depleted sample was below the detection limit of Qubit. Library construction using Rapid PCR Barcoding Kit with host-depleted gDNA samples yielded 63ng/ul library DNA. Host-depletion resulted in a significant reduction in the proportion of human reads from 97% to 14%, while simultaneously enriching microbial reads from 0.34% to 85.96%.

Conclusions ONT sequencing can be adopted for both WGS-based and 16S+ITS-based metagenomics NGS applications. DNA from host depleted samples can be sequenced on ONT by incorporating PCR amplification during library preparation.

Engaging Multiple Myeloma, Colorectal Cancer, and Cholangiocarcinoma Patients in Cancer Disparity Research

POSTER 202

Reyka G. Jayasinghe[1], Fernanda M. Rodrigues[1], Robert J. Mashl[1], Song Cao[1], Yizhe Song[1], Chien-Wei Peng[1], Erin Linnenbringer[2], Bailey Martin-Giacalone[2], Erik Storrs[1], Fankun Zeng[1], Yingduo Song[1], Hasani Bland[3], Danielle Grasso[3], Mark Fiala[4], Ravi Vij[4], Robert S. Fulton[5], David H. Spencer[6], Eric J. Duncavage[7], Bettina F. Drake[2], Ryan C. Fields[8], Graham A. Colditz[9], Li Ding[1]

1 Division of Oncology, Department of Internal Medicine, Washington University School of Medicine, St. Louis, MO
2 Department of Surgery, Division of Public Health Sciences, Washington University School of Medicine, St. Louis, MO
3 Department of Surgery, Washington University School of Medicine, St. Louis, MO
4 Bone Marrow Transplantation & Leukemia Section, Division of Oncology, Washington University School of Medicine, St. Louis, MO
5 Department of Genetics, McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO
6 Department of Pathology, Washington University School of Medicine, St. Louis, MO
7 Division of Anatomic and Molecular Pathology, Department of Pathology and Immunology, Washington University School of Medicine in Saint Louis, MO
8 Section of Hepatobiliary Surgery, Department of Surgery, Washington University School of Medicine, St. Louis, MO
9 Division of Public Health Sciences, Washington University School of Medicine, St. Louis, MO

To engage diverse populations in genomics research and better characterize the genomic and transcriptional cancer landscape of understudied ethnic groups, Washington University is participating in the NCI Participant Engagement and Cancer Genome Sequencing (PE-CGS) Network to sequence underrepresented groups in Colorectal cancer (CRC), Multiple Myeloma (MM) and Cholangiocarcinoma (CHOL). More specifically, our center is actively recruiting patients with CHOL, Black Americans under age 65 with CRC and Black Americans with MM as these malignancies disproportionally affect the Black community. Due to the rarity and aggressive nature of Cholangiocarcinoma (CHOL), all patients are eligible as there is a need to understand the underlying genetics of the disease. Under the PE-CGS effort we engage with these communities, conduct comprehensive genomic testing, and provide findings back to participants, all while gaining valuable insight into how these communities engage with genetic research and results.

(continue to page 37)

(continued from page 36)

The genomic characterization unit (GCU), which oversees sequencing for the study, has developed an automated pipeline to analyze Whole Exome Sequencing (WES) and RNA Sequencing data derived from paired tumor/normal samples and developed a prioritization scheme for reporting somatic and germline variants. Further, our reporting incorporates variant classification matches to currently recruiting and anticipated future Clinical Trials (NIH). In addition to the genomic findings, we have also performed other research-based assays including Xenium and single-cell RNA-sequencing to evaluate the transcriptional and spatial architecture of a subset of tumors. Our research will provide novel findings to the research community related to these under-characterized tumors and patient groups.

Enhancing custom large hybridization panels with GenScript's GenTitan HD 8M oligo pool

POSTER 203

Xin Ray Tee[1], Xiaowei Yan[1], Cedric Wu[1], Vincent Magrini[1]

1 GenScript USA Inc., Innovation Center, Seattle, WA

The adoption of custom large hybridization panels in clinical settings faces enduring challenges, encompassing cost, probe uniformity, and turnaround times. Addressing these hurdles, GenScript introduces the GenTitan HD 8M Oligo Pool. Leveraging cutting-edge CMOS technology, we excel in synthesizing high-fidelity oligo precursors at an unprecedented scale, accommodating up to 32 million oligos simultaneously. This breakthrough empowers clinical researchers with unparalleled capabilities in gene assembly, mutant library construction, and hybridization probe preparation.

GenTitan Probes demonstrate competitive on-target rates across various panel sizes, from 940kb panels (GenScript @ 88±2% vs Competitor @ 85±2%), to 13.9Mb panels (GenScript @ 93±2% vs Competitor @ 93±1%) and more. These probes can detect variant allele fractions as low as 0.5%, including precise identification of copy-number alterations at a 2-fold difference.

A hallmark feature of the GenTitan HD 8M Oligo Pool is flexibility, enabling users to tailor probe levels for optimized capture efficiency. This capability reduces overabundant sequences and enhances coverage in challenging genomic regions. Our proprietary 'rebalancing' algorithm further augments this capability, ensuring uniform coverage across regions of interest by considering factors like GC content, tandem repeats, masked regions, and mappability.

Illustrated in a case study with a 4.56Mb panel design, initial captures revealed significant coverage disparities, particularly in AT-rich regions. Through algorithmic rebalancing, we achieved substantial improvement in coverage uniformity (n=4; F80P 1.62 →1.33, t-test, p=5×10-4) while maintaining competitive on-target capture rates (Original @ 89±2%; Rebalanced @ 90±3%). Targets captured within a 2-fold range of the median increased from 95% to 98%. Notably, the rebalanced pool managed to reclaim 60% of AT dropouts noted in the original design.

Looking ahead, we are integrating AI-driven frameworks into our rebalancing algorithms to enhance predictive accuracy and streamline panel design. This evolution promises efficient and consistent panel captures, potentially eliminating the need for rebalancing.

Join us in revolutionizing genomics with collaborative services that deliver precise, standardized panel designs for accelerated clinical research in hybrid-target enrichment. Welcome to the future of precision health with GenScript.

EpiPlex: a platform for the multiplexed detection of RNA modifications

POSTER 204

Andrew Price[1], Zachary Miles[1], Zac Johnson[1], Byron W. Purse[1], Gudrun Stengel[1]

1 Alida Biosciences, San Diego, CA

Here, we present AlidaBio's progress in developing and commercializing EpiPlex™, a high-throughput and multiplexed RNA modification profiling technology. The comprehensive EpiPlex™ platform includes a library preparation kit and a suite of machine learning-powered bioinformatic tools, offering genomic researchers a unique layer of multiomic insight that is scalable and accessible.

The epitranscriptome is the entirety of naturally occurring chemical modifications to RNA. These modifications are involved in regulating seemingly all aspects of RNA metabolism, including splicing, translation initiation and fidelity, trafficking, RNA folding and stability, and transcript lifetime. Although more than 170 different RNA modifications have been identified, sequence resolved methods that can detect even a few RNA modifications simultaneously are lacking. Multiplexed detection of RNA modifications is significant as mass spectrometry studies show complex and correlated changes of the epitranscriptome depending on cell or disease state. Furthermore, the precision health applications of the EpiPlex assay are profound, as it enables the characterization of disease-specific RNA modification patterns, paving the way for personalized therapeutic strategies and early diagnosis.

The EpiPlex assay is a single day proximity barcoding assay with sequencing readout that detects multiple targets, while providing relative quantification and accommodating low RNA input, making it compatible with clinically relevant samples. The current version of the assay detects m6A and inosine using protein binder-directed barcoding, and is extendable to additional modifications. In addition to providing a technological overview and validation of assay accuracy with RNA modification writer knockout cell lines we will present biological data showing aberrant regulation of m6A and inosine levels in disease/normal FFPE sample pairs. Data shown will feature EpiScout™ modification site identification and quantitation which leverage machine learning algorithms and a novel system of spike-in assay controls. The excellent sensitivity and quantitative capabilities of the EpiPlex platform has the potential to provide new insight into the molecular pathology of disease states.

Estimating the prevalence of dominant Mendelian disorders using the gnomAD database

POSTER 205

Mutaz Amin[1,2], Samantha Baxter[1], Kaitlin Samocha[1,3], Anne O'Donnell-Luria[1,2,3]

1 Broad Institute of MIT and Harvard, Cambridge, MA
2 Boston Children's Hospital, Harvard Medical School, Boston, MA
3 Massachusetts General Hospital, Harvard Medical School, Boston, MA

Background: The Genome Aggregation Database (gnomAD) is an extensive genomic sequencing dataset that compiles genetic variations from over 700,000 individuals. Using population databases like gnomAD to estimate the genetic prevalence of monogenic disorders has proven effective for autosomal recessive disorders. However, dominant and X-linked monogenic disorders, especially those associated with severe or early onset phenotypes, pose a challenge because gnomAD is likely to be depleted for variants that affect reproductive fitness.
Methods: We developed a bimodal approach to estimate the prevalence of autosomal dominant monogenic disorders. This method uses a mutation rate model with the observed/expected (o/e) ratio of missense and predicted loss-of-function variants from gnomAD, which is most relevant for de novo dominant conditions, alongside the gnomAD allele counts, pertinent to later-onset inherited conditions. We use the average o/e ratio of missense and loss-of-function variants for non-disease-causing genes as minuend for unobserved potential disease-causing variants. Additionally, we created a lay-language point-system survey, whose results will guide users in determining whether to apply the mutation rate model or the allele count model.
Results: The bimodal method is currently being evaluated for a subset of dominant monogenic disorders in collaboration with patient advocacy organizations in partnership with Chan Zuckerberg Initiative's Rare As One Network. This method is being compared to existing prevalence data from Orphanet and other published articles. We identified pitfalls and causes of less reliable estimates, such as gain-of-function rather than loss-of-function disease mechanism and developed a specific FAQs page to address these challenges.
Conclusion: While this work is preliminary, understanding the structure of the gnomAD database and interpreting rare observed and unobserved variants from gnomAD are essential steps in estimating the prevalence of dominant monogenic disorders.

From targeted panels to whole genome sequencing: analytical opportunities, challenges, and case insights in clinical germline disease diagnosis

POSTER 206

Xuanjin Cheng[1], Karen L. Mungall[1], Stephen Yip[2], Steven J.M. **Jones**[1]

1 Canada's Michael Smith Genome Sciences Centre, Vancouver, British Columbia, Canada
2 BC Cancer Agency, Surrey, British Columbia, Canada

Advancing genomic diagnostics, our clinical services are transitioning from targeted panels to whole genome sequencing (WGS) for germline disease diagnosis, following best practices by the Medical Genome Initiative and the College of American Pathologists (CAP) guidelines. This shift promises comprehensive variant detection and deeper genetic insights, while also presenting analytical challenges and opportunities. This presentation compares the performance of targeted panels and WGS in detecting small variants and copy number variations (CNVs). Additionally, we discuss strategies to mitigate technical challenges in next-generation sequencing (NGS) and approaches to minimize false negatives while controlling false positives through case studies. Our goal is to enhance clinical utility by leveraging the full potential of NGS-based genomic assays.

Genomic testing in neonatal hypotonia: an international consortium experience

Sarah U. Morton[1,2,3,4], Gregory Costain[5,6,7,8], Courtney E. French[9], Emma Wakeling[10], Anna Szuto[8,11], John Christodoulou[12,13], Ronald D. Cohn[5,6,7], Basil T. Darras[14], Monica H. Wojcik[1,2,3,4,9], Alissa M. D'Gama[1,2,15], IPCHiP Executive Committee, James J. Dowling[6,7,8,16], Sebastian Lunke[12,17], Francesco Muntoni[10,18], F Lucy Raymond[19], David Rowitch[19,20], Alan H. Beggs[2,3,4,9], Zornitza Stark[12,21], Pankaj B. Agrawal[2,3,4,9,22]

1 Division of Newborn Medicine, Boston Children's Hospital, Boston, MA
2 Department of Pediatrics, Harvard Medical School, Boston, MA
3 The Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, MA
4 The Broad Institute of MIT and Harvard, Cambridge, MA
5 Division of Clinical & Metabolic Genetics, The Hospital for Sick Children, Toronto, ON, Canada
6 Program in Genetics & Genome Biology, SickKids Research Institute, Toronto, Ontario, Canada
7 Department of Paediatrics, University of Toronto, Toronto, Ontario, Canada
8 Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada
9 Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA
10 North East Thames Regional Genetic Service, Great Ormond Street Hospital Trust, London, United Kingdom
11 Department of Genetic Counselling, The Hospital for Sick Children, Toronto, Ontario, Canada
12 Murdoch Children's Research Institute and Department of Paediatrics, University of Melbourne, Melbourne, Victoria, Australia
13 Discipline of Child and Adolescent Health, Sydney Medical School, University of Sydney, Sydney, New South Wales, Australia
14 Department of Neurology, Boston Children's Hospital, Boston, MA
15 Epilepsy Genetics Program, Department of Neurology, Boston Children's Hospital, Boston, MA
16 Division of Neurology, The Hospital for Sick Children, Toronto, Ontario, Canada
17 Department of Pathology, University of Melbourne, Melbourne, Australia
18 National Institute for Health Research Great Ormond Street Hospital Biomedical Research Centre, Great Ormond Street Institute of Child Health, University College London, London, United Kingdom
19 Departments of Medical Genetics and Paediatrics, University of Cambridge, Cambridge, United Kingdom
20 Division of Neonatology, Department of Pediatrics, UCSF, San Francisco, CA
21 Australian Genomics Health Alliance, Parkville, Australia
22 Division of Neonatology, Department of Pediatrics, University of Miami and Holtz Children's Hospital, Jackson Health System, Miami, FL

(continue to page 43)

(continued from page 42)

Hypotonia is a common presentation among infants in the neonatal intensive care unit (NICU). We completed a retrospective study of infants with hypotonia to determine the yield of exome/genome sequencing (ES/GS) across five major pediatric tertiary hospitals in US, Canada, UK and Australia. All infants who were cared for in the NICU and underwent genetic evaluation were included. We identified 147 infants with unexplained hypotonia, 80 (54%) of whom had hypotonia as the primary finding, while 67 (46%) had additional multisystemic involvement. Seventy-five (51%) underwent ES, 44 (30%) GS, two (1%) both ES and GS, and 26 (18%) were admitted before ES/GS became available. The majority of ES/GS was done in a rapid fashion with turnaround time of less than two weeks. Of the 121 infants who underwent ES and/or GS, 72 (60%) received a molecular diagnosis. In addition, two infants with mitochondrial genome variants were diagnosed by mitochondrial genome sequencing after negative ES, and one infant needed targeted testing to identify a short tandem repeat expansion missed by GS. The proportion diagnosed by ES/GS was not significantly different between infants with hypotonia as the primary finding (37/56, 66%) and infants with multisystemic symptoms (35/65, 58%, p-value=0.20). In infants receiving a genetic diagnosis, testing was more likely to have an impact on care (57/66 vs 14/33, p-value=1.0E-05). The most common impacts on care of infants who received a molecular diagnosis, as identified by the clinical team, were improving medical care plans such as providing earlier gastrostomy tube placement in anticipation of poor oral feeding (22), adjusting clinical goals of care (20), and offering new treatments/evaluation (5). In conclusion, rapid ES/GS has a high diagnostic yield in infants with hypotonia and should be considered as a first-line genetic test for infants with unexplained hypotonia, given the capacity of early diagnosis to greatly influence management.

Germline predisposition in multiple myeloma

POSTER 207

Fernanda Martins Rodrigues[1,2,3], Jagoda Jasielec[4], Melody Perpich[4], Aelin Kim[4], Luke Moma[4], Erik Storrs[1,2,3], Yize Li[1,2,3], Michael C. Wendl[1,2,3], Reyka G. Jayasinghe[1,2,3], Mark Fiala[1,2], Andrew Stefka[4], Benjamin Derman[4], Andrzej J. Jakubowiak[4], John F. DiPersio[1,2,5], Ravi Vij[1,2,5], Lucy A. Godley[6], Li Ding* [1,2,3,5,7]

1 Washington University School of Medicine, Department of Medicine, St. Louis, MO
2 Washington University School of Medicine, Division of Oncology, St. Louis, MO
3 Washington University School of Medicine, McDonnell Genome Institute, St. Louis, MO
4 The University of Chicago, Section of Hematology/Oncology, Department of Medicine, Chicago, IL
5 Washington University School of Medicine, Siteman Cancer Center, St. Louis, MO
6 Northwestern University, Hematology Oncology Division, Evanston, IL
7 Washington University School of Medicine, Department of Genetics, St. Louis, MO

Multiple myeloma (MM) is an incurable malignancy, characterized by clonal proliferation of malignant plasma cells in the bone marrow, monoclonal immunoglobulins in urine and serum, and organ dysfunction. It is often preceded by a premalignant condition called monoclonal gammopathy of undetermined significance (MGUS), which progresses to asymptomatic smoldering MM (SMM), and ultimately, MM. Risk factors associated with MM include age, male gender, race/ethnicity, family history, and history of MGUS, suggesting a germline contribution to MGUS and MM.

Although considerable progress has been made in cataloging somatic events underlying MM development and progression, little is known about its genetic predisposition. The CoMMpass study by MMRF (Multiple Myeloma Research Foundation) is a longitudinal, prospective observational study involving the collection and analysis of sequencing and clinical data from >1K MM patients at diagnosis and relapse. Here, we present the first combined cohort and family study of rare germline predisposition variants (GPVs) in 954 unrelated individuals with MM from the MMRF and 82 MM families. We identified pathogenic/likely pathogenic variants across all age groups in 9.1% of sporadic and 18% of familial cases. Implicated genes include genes previously suggested to have an association with MM risk (DIS3, KDM1A, USP45); genes involved in predisposition to other cancer types but not previously associated with MM predisposition (ATM, BRCA1/2, CHEK2, PMS2, POT1, PRF1, TP53); and BRIP1, EP300, and FANCM in individuals of African ancestry. Variants were characterized using loss of heterozygosity (LOH), biallelic events, and gene expression analyses, revealing 31 variants in 3.25% of sporadic cases for which pathogenicity was supported by multiple lines of evidence. Our results suggest that disruption of DNA damage repair related pathways may play a role in MM susceptibility.

(continue to page 45)

(continued from page 44)

In summary, the results and analysis strategies used here will allow for efficient and cost-effective discovery of genetic changes relevant to MM etiology ultimately impacting our overall understanding of the genetics underlying MM predisposition and improving surveillance in high-risk groups and potential personalized therapeutic strategies.

Health Metadata Commons: unifying biobank and omics metadata

POSTER 208

Zsuzsanna Hollander[1], Annaïck Carles[1], Mohammed Hasham[1], Federica di Palma[1]

1 Genome British Columbia, Vancouver, British Columbia, Canada

The rapid advancement of omics research has generated a vast amount of data, creating a critical need for effective data management and sharing strategies. Furthermore, numerous biobanks exist that house extensive collections of human samples. However, today's data silos create barriers to information sharing and collaboration. The Health Metadata Commons addresses this need by providing a centralized platform that facilitates the access, contribution, and integration of comprehensive genomic research metadata and biobank data. Designed to support researchers across various disciplines, the Health Metadata Commons aims to enhance data discoverability and reuse, thereby accelerating scientific discoveries and innovations.

Genome British Columbia developed the Health Metadata Commons, a cloud-based solution using the Oracle APEX technology platform. This platform maintains an inventory of omics datasets as well as biological samples that hold the potential for future profiling. One of the core features of the Metadata Commons is its user-friendly interface, which allows researchers to easily submit, search, and retrieve metadata. The metadata collected includes basic information about the project and aggregate information about the patient cohort, sample, and omics type developed. It employs standardized metadata schemas and ontologies to ensure consistency and compatibility across projects. By leveraging advanced data harmonization techniques, the platform integrates metadata from multiple omics types, making it easier for researchers to locate relevant data and conduct meta-analyses, including multi-omics. In addition, researchers can save their preferences for specific disease conditions, omics types to get notified by email when projects satisfying their criteria are submitted to the platform.

The Health Metadata Commons holds promise for stretching the research funds further, leading to new findings and discoveries by reusing data beyond its initial purpose. As a funding agency focused on applying genomics to pressing societal and economic challenges, Genome British Columbia developed the Health Metadata Commons to foster a collaborative culture and provide more equitable and integrated metadata to the research community worldwide.

How the nation's largest integrated healthcare system is accelerating precision medicine initiatives: Veterans Health Administration (VHA) innovation ecosystem

POSTER 209

R. Relova[1], A. Callahan[2], S. Patel[3], P. Joshi[4], A. Goetschius[2], A. Gant-Curtis[2], K. Teague[2], C. Zanetti[3]

1 United States Department of Veterans Affairs (VA), Palo Alto, CA
2 United States Department of Veterans Affairs (VA), Washington, DC
3 United States Department of Veterans Affairs (VA), Charleston, SC
4 United States Department of Veterans Affairs (VA), San Francisco, CA

Background: Precision Medicine (PM) has emerged in recent years as a potential way to deliver personalized, comprehensive care to patients with a wide range of diseases. PM enables diagnosis of disease, treatment therapies, and prevention to be individualized, proactive, predictive, and precise. However, several challenges persist when it comes to incorporating these strategies in routine clinical care. Large, complex datasets combined with inadequate technology, and interoperability of healthcare systems have consistently been hurdles in implementing precision care. The VHA Innovation Ecosystem identifies and supports promising evidence-based practices and innovative solutions proposed by frontline healthcare workers for implementation and sustainment in VHA medical facilities or service networks.
Purpose: VHA Innovation Ecosystem aims to facilitate the adoption of Precision Medicine's innovative approach in care systemwide by supporting the implementation of Pharmacogenomics (PGx) which tailors a patient's prescription based on one's genetic profile.
Current State: VHA is dedicated to continually innovating its systems, technology, and practices to provide high-quality care to US Veterans. Pharmacogenomics (PGx) serves as an essential clinical decision support (CDS) tool for clinicians at point of prescribing. Realizing the impact of PM initiatives on health outcomes, VHA Innovation Ecosystem is providing necessary resources and staffing support to leverage a platform called Data Arch that enables the clinicians' timely access to PGx-guided medication and dosing recommendations. PGx is pivotal in alleviating unintended side effects which can result from the current one-size-fits-all approach. This PM approach will also reduce the cost of treatment by eliminating ineffective therapeutic plans at outset solely based on data insights.
To date, VHA has over 7,000 unique clinicians actively ordering PGx tests averaging over 4,800 PGx reports/month across 140 VHA medical facilities.
Conclusion: Developing an Innovation infrastructure serves as a promising model for other healthcare systems seeking to accelerate the spread and adoption of novel clinical initiatives. The VHA Innovation Ecosystem facilitates the seamless integration of Precision Medicine interventions in clinical settings.

Impact of Missense Mutations on GPx7 and GPx8: Insights into Protein Stability, Pathogenesis, and Precision Health Applications

POSTER 210

Shaolei Teng[1], Adebiyi Sobitan[1]

1 Howard University, Washington, DC

Somatic and genetic mutations of GPx7 and GPx8 have been associated with intellectual disability, microencephaly, and tumors in different parts of the body. Evolving latest among the GPx enzymes and located in the endoplasmic reticulum, GPx7 and GPx8 lack the glutathione (GSH) binding domain but possess peroxidase activity that aids the body in responding to cellular stress. Despite their importance, these enzymes remain understudied. This study aims to elucidate the structural and stability consequences of missense mutations in GPx7 and GPx8, providing insights into the pathogenic mechanisms involved in different cancer types. This research could inform clinical diagnosis, treatment strategies, and the development of targeted therapies, contributing to Precision Health. We performed saturated computational mutagenesis to analyze 2926 and 3971 missense mutations of GPx7 and GPx8, respectively. For GPx7, G153H and G153F were highly destabilizing, while E93M and W142F were highly stabilizing. For GPx8, N74W and G173W caused the most instability, whereas S70I and S119P increased stability. Our analysis indicates that highly destabilizing mutations are more likely pathogenic compared to highly stabilizing mutations. The predictions from this study show a significantly high correlation (p-value < 2e-16) with predictions from Alphamissense and Meta-SNP. In conclusion, the computational mutagenesis of missense mutations in GPx7 and GPx8 offers valuable insights into their impact on protein structure and stability, contributing to our understanding of pathogenesis and cancer progression, and supporting the advancement of precision health strategies.

Implementing ultra-sensitive, ctDNA-based liquid biopsy for molecular monitoring in high-risk pediatric tumors

POSTER 211

Mark J. Cowley[1,2], Rob Salomon[1,*], Wenhan Chen[1,2,*], Charles Shale[3,*], Mojgan Toumari[1], Aileen Lowe[1], Wenyan Li[1], Jingwei Chen[1,2], Sajad Razavi-Bazaz[1], Paulette Barahona[1], Louise Cui[1], Chelsea Mayoh[1,2], Sam El-Kamand[1], Marie Wong[1,2], Loretta MS Lau[1,2,5], Noemi Fuentes-Bolanos[1,2,5], Edwin Cuppen6, Dong-Anh Khuong-Quang[7,8], Jordan Staunton[1,2,5], Marion K Mateos[1,2,5], Toby Trahair[1,2,5], Vanessa Tyrrell[1,2], David S. Ziegler[1,2,5], Michelle Haber[1,2], Paul G. Ekert[1,2,8,9], Peter Priestley[3]

1 Children's Cancer Institute, Lowy Cancer Research Centre, UNSW, Sydney, Australia
2 School of Clinical Medicine, UNSW Medicine & Health, UNSW, Sydney, Australia
3 Hartwig Medical Foundation Australia, Sydney, New South Wales, Australia
4 Discipline of Paediatrics and Child Health, School of Clinical Medicine, UNSW Medicine & Health, UNSW Sydney. Australia
5 Kids Cancer Centre, Sydney Children's Hospital Randwick, Sydney, Australia
6 Hartwig Medical Foundation, Amsterdam, The Netherlands
7 Children's Cancer Centre, Royal Children's Hospital, Melbourne, Victoria, Australia
8 Department of Paediatrics, Murdoch Children's Research Institute, University of Melbourne, Melbourne, Victoria, Australia
9 Cancer Immunology Program, Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia
* These authors contributed equally

We assessed the clinical utility of high-sensitivity ctDNA-based liquid biopsy in children with diverse high-risk cancers, including brain, hematological, sarcoma, neuroblastoma, and extra-cranial solid tumors. We hypothesized that personalized assays, informed by comprehensive genomic profiling of the tumor, could enhance diagnosis and disease progression insights.
We retrospectively analyzed 169 patients with 661 blood or CSF samples from the ZERO Childhood Cancer Precision Medicine program (Australia). Using whole genome sequencing of tumour biopsy and matched germline, we identified high-confidence, clonal somatic mutations. We designed personalized ctDNA capture panels covering ~300 patient-specific mutations and a fixed backbone of ~100 hotspot and treatment resistance mutations. These were applied across multiple samples taken at different timepoints for the same patient and sequenced at >10,000x depth with duplex error correction. We developed novel informatic pipelines (redux and wisp) to reduce noise, improve sensitivity, ensure accurate disease detection even at very low tumor burdens, and provide tumor fraction estimations for tracking disease over time.
We achieved an average detection limit between 10^-4 and 10^-5 ctDNA fraction, overcoming limitations of low DNA input and few recurrent mutations.

(continue to page 50)

(continued from page 49)

We observed a wide range of ctDNA burden, from 0.001% to 100%. Our approach identifies ctDNA in most samples collected when disease is present, based on medical imaging. For CNS tumors, ctDNA detection in blood is lower than CSF, but still significantly higher compared to similar studies. We identified residual disease in some patients who appeared to have completely responded to treatment according to radiological imaging and could identify disease relapse 1–6 months before clinical manifestation in some patients. We also identified novel variants emerging during disease progression through analysis of the fixed hotspot mutations.
In conclusion, we developed an ultra-sensitive, tumor-informed, personalized ctDNA-based approach to diverse high-risk childhood cancers. This improved sensitivity enhances longitudinal disease tracking and allows earlier relapse detection. The ability to detect the emergence of targetable variants not found in the original tumor suggests this approach could identify additional treatment options that arise with disease evolution.

In silico genetic screening to identify cancer cell vulnerabilities conferred by tumor suppressor gene mutations

POSTER 212

Yuka Takemon[1,2], Erin D. Pleasance[2], Veronika Csizmok[2], Andrew J. Mungall[2], Richard A. Moore[2], Eric Chuah[2], Karen L. Mungall[2], Eleanor Lewis[2], Howard J. Lim[3], Daniel J. Renouf[3,4], Steven J.M. Jones[2,5], Janessa Laskin[4], Marco A. Marra[1,2,5]

1 Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia, Canada
2 Canada's Michael Smith Genome Sciences Centre, BC Cancer Research Institute, Vancouver, British Columbia, Canada
3 Department of Medical Oncology, BC Cancer, Vancouver, British Columbia, Canada
4 Pancreas Centre BC, Vancouver, British Columbia, Canada
5 Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada

Tumour suppressor genes have the highest mutational frequencies across most cancer types and are notoriously ineffective therapeutic targets. Here, we use KMT2D, a tumour suppressor gene, as a case study to demonstrate how in silico mapping of genetic interaction networks can be used to characterize KMT2D function and identify vulnerabilities that can be leveraged for therapeutic intervention. KMT2D encodes a histone lysine methyltransferase and is a member of the chromatin-remodeling COMPASS complex. Somatic KMT2D alterations have been found in several cancer types, including follicular lymphomas (89%), diffuse large B-cell lymphomas (32%), urothelial bladder carcinomas (27%), and gastric cancers (12%). KMT2D alterations and/or loss of activity have been associated with inferior survival outcomes in several cancer types, promoting tumorigenesis, cancer recurrence, transcriptional stress, and genomic instability. We mapped KMT2D genetic interaction networks using cell lines representing nine cancer types with frequent KMT2D alterations. We identified several candidate interactors, including synthetic lethal interactions with WRN, MDM2, NDUFB4, and TUBA1B, which encode targets of existing and in-development therapeutics, making them potentially viable drug targets in cancers harbouring KMT2DLOF alterations. We showed that KMT2DLOF cancer cell lines with microsatellite instability (MSI) have aberrant microsatellite expansion, a feature attributed to WRN perturbation vulnerability. We also used The Cancer Genome Atlas COAD-READ cohorts and observed significantly higher tumour mutational burden (TMB), a marker for immune checkpoint inhibitors (ICIs), in KMT2D-deficient cases when compared with KMT2D-proficient cases. We further confirmed these observations using data from the Personalized Oncogenomics Program (POG; NCT02155621), a BC Cancer precision cancer genomic medicine study. Altogether, we identified four targetable candidate lethal interactors and observed high TMB in KMT2D-deficient cancers in several cohorts, which may provide a rationale for investigating ICI treatments in such cases. Our work thus illustrates a framework for identifying targetable vulnerabilities in cancers driven by “undruggable” genetic alterations.

Investigating the mechanisms of incomplete penetrance in a rare disease cohort

POSTER 301

Mugdha Singh[1,2], Emily Groopman[1,2], Sarah Stenton[1,2], Sanna Gudmundsson[1,2,3], Anne O'Donnell-Luria[1,2]

1 Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA
2 Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA
3 Science for Life Laboratory, Department of Gene Technology, KTH Royal Institute of Technology, Stockholm, Sweden

Variant interpretation remains challenging and can limit rare disease diagnosis. One factor complicating variant interpretation is incomplete penetrance, a phenomenon where only some individuals who carry a pathogenic variant will develop the disease. Diagnosis in these cases is restricted by our limited understanding of how genomic factors modify the phenotypic outcomes of incompletely penetrant pathogenic variants. The primary objective of this project is to understand the molecular mechanisms of incomplete penetrance that result in the absence of disease symptoms in unaffected carrier parents. We are undertaking a large study of incomplete penetrance using exome and/or genome sequencing data of 4,389 families from the Broad Institute Center for Mendelian Genomics (CMG), a research center of the NHGRI GREGoR consortium. We focused on variants in genes associated with severe, early-onset, dominant disorders and developed an approach to identify inherited (from an unaffected parent) putative pathogenic, incompletely penetrant variants. Our systematic approach focused on ClinVar pathogenic (P) and likely pathogenic (LP) variants as well as high confidence predicted loss of function (pLoF) variants in haploinsufficient disease genes. We identified 1,148 (P/LP/pLoF) potentially pathogenic variants. Further assessment of inheritance pattern, evidence of pathogenicity, and phenotype overlap refined the list to 199 potentially pathogenic variants from 165 families. We hypothesize that cis-regulatory variants can modify phenotypes in some cases of incomplete penetrance. We will investigate the possible rescue mechanisms in unaffected parents compared to affected probands through molecular characterization of the genome sequence, transcriptomics, methylation, and/or proteomics profiles. By understanding the molecular determinants, we aim to increase diagnostic yield, to better understand the underlying mechanism of incomplete penetrance, and to use this understanding to devise strategies for therapeutic intervention for these disorders.

Investigating the variant-specific phenotypes of the GABA transporter solute carrier 6 member 1 (SLC6A1) in Mendelian disease traits

POSTER 302

Nikhita Gogate[1], Kristy L. Jay[1,2], Ryan German[1,2], Vanessa Gomez[1,2], National Brain Gene Registry, Michael F. Wangler[1,2], Jennifer E. Posey[1]

1 Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX
2 Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX

Rare heterozygous pathogenic variants in solute carrier 6 member 1 (SLC6A1) are associated with Myoclonic-atonic Epilepsy (MIM#: 616421). SLC6A1 encodes the sodium- and chloride-dependent GABA transporter 1 (GAT-1), which plays a crucial role in maintaining synaptic neurotransmitter balance. Individuals with damaging variants in SLC6A1 exhibit epilepsy and developmental delay as prominent phenotypes. In addition, features such as autistic traits, intellectual disability, hypotonia, sleep disorders, and attention-deficit/hyperactivity disorder may be observed with variable severity. The current lack of comprehensive data on the phenotypic landscape of SLC6A1-neurodevelopmental disorders limits expectant management and effective treatment development. In this study, we recruited 14 patients with SLC6A1-related disorders through the National Brain Gene Registry to explore symptom variability by variant type. Deep clinical phenotyping highlights neurological abnormalities and cognition impairment to be highly impacted. Additionally, the rapid neurodevelopmental assessment protocol (RNAP) highlighted that cognition and communication were the most affected domains, significantly impacting quality of life. Seizures are managed with general anti-epileptics with varying levels of success; however, cognitive impairment persists regardless of seizure suppression. As a result, there is uncertainty regarding the biological underpinnings of the neurodevelopmental phenotypes. We utilize Human Phenotype Ontology (HPO)-based phenotypic clustering analysis for generating new hypotheses regarding the involvement of SLC6A1 variants in neurodevelopmental phenotypes and uncover phenotypic subgroups that may be driven by distinct variants or variant types in SLC6A1. Even with highly similar phenotypic scores associated with SLC6A1 variant alleles, subclusters of related phenotypic profiles can be identified. We observe that SLC6A1 loss-of-function traits cluster together with similar phenotypes, however missense variants are not associated with a strong genotype-phenotype correlation. We support this finding with in vivo evaluation in a Drosophila model. Thus, investigating the unique spectrum of clinical phenotypes of SLC6A1-related disorders, we can begin to understand the genotype-phenotype associations underlying the variability in disease expression. These insights will inform clinical diagnosis, prognosis, treatment planning, and the design of targeted therapeutics.

Abstract Selected Talk

Knowledge informed machine learning to identify undiagnosed patients with rare diseases in electronic health records

Theodore Morley[1,2], Lucas Richter[1,2], Jon Morra[3], Lisa Bastarache[2,4], Douglas Ruderfer[1,2,4]

1 Vanderbilt University Medical Center, Division of Genetic Medicine, Nashville, TN
2 Vanderbilt University Medical Center, Center for Digital Genomic Medicine, Nashville, TN
3 Zefr, Los Angeles, CA
4 Vanderbilt University Medical Center, Department of Biomedical Informatics, Nashville, TN

Introduction: Despite comprising a significant proportion of the population, patients with rare genetic diseases endure extended diagnostic odysseys. The many thousands of genetic diseases challenge the recognition necessary to shorten time to diagnosis. However, substantial descriptive information on these diseases exists and with more patients receiving genetic testing as part of routine healthcare, there is an opportunity to develop algorithmic approaches using both knowledge and patient data to aid recognition and diagnosis.
Methods: Our cohort consisted of 5,279 patients diagnosed with 949 distinct diseases from the electronic health record (EHR). We mapped Online Mendelian Inheritance in Man (OMIM) descriptions of 7,411 diseases into PheWAS codes, consistent with the EHR data. Patient EHR records were censored to the year prior to testing or diagnosis to prevent data leakage. For each patient, we combined codes from the EHR with codes for a given OMIM disease description to generate a patient by disease representation. This representation was considered a case only when the patient had that specific genetic disease and a control otherwise including all diseases for patients without a diagnosis. Multiple classification methods were evaluated through 5-fold cross validation, and the best model was selected based on average precision in the validation set.
Results: Despite the median disease category having only a single case, we achieve a mean per-disease AUROC of 0.77, median per-disease AUROC of 0.85. This performance holds even in 103 disease categories which were not present in the training set, which have a mean per-disease AUROC of 0.8, median per-disease AUROC of .9. Top 20 recall across all patients (the proportion of patients whose true disease is within the top 20 highest ranked diseases for them) is 27%.
Discussion: Our results demonstrate the power of integrating amassed knowledge of rare genetic diseases with data collected at scale from EHRs. This is evident by the prediction performance of diseases that were never seen in our training set. Using only publicly available resources and diagnostic codes, our approach enables a portable means to identify and suggest diagnoses of patients with rare genetic diseases which if implemented may shorten diagnostic odysseys.

Abstract
Selected
Talk

Large-scale discovery of gene regulatory elements for cis-regulation therapy for autism spectrum disorder and neurodevelopmental delay

Nicholas F. Page[1,2,3], Troy A. McDiarmid4, Florence M. Chardon[4,5], Stephan J. Sanders[1,3,6], Jay Shendure[4,5,7,8,9], Nadav Ahituv[2,3]

1 Department of Psychiatry and Behavioral Sciences, Kavli Institute for Fundamental Neuroscience, Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA
2 Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA
3 Institute for Human Genetics, University of California, San Francisco, San Francisco, CA
4 Department of Genome Sciences, University of Washington, Seattle, WA
5 Seattle Hub for Synthetic Biology, Seattle, WA
6 Institute of Developmental and Regenerative Medicine, Department of Paediatrics, University of Oxford, Oxford, United Kingdom
7 Brotman Baty Institute for Precision Medicine, Seattle, WA
8 Howard Hughes Medical Institute, Seattle, WA
9 Allen Discovery Center for Cell Lineage Tracing, Seattle, WA

Haploinsufficiency, where only one of the two gene copies is functional, is a major risk factor for autism spectrum disorder (ASD) and neurodevelopmental delay (NDD). Cis-regulation therapy (CRT) is a promising therapeutic approach to rescue haploinsufficiency, combining a nuclease deficient Cas9 (dCas9) fused to a transcriptional activator that targets via an sgRNA the regulatory element/s of the haploinsufficient gene to upregulate the expression levels of the remaining functional gene copy. However, current methods to identify and optimize therapeutic sgRNAs are based on a 'one-by-one' manner, which is both slow and costly. To identify ASD/NDD regulatory elements and accelerate the discovery of sgRNAs that could be used for CRT, we utilized both massively parallel reporter assays (MPRAs) and single-cell RNA-seq combined with multiplex CRISPR activation (CRISPRa) in iPSC-derived excitatory neurons. We initially annotated 5,425 candidate brain enhancers for 337 ASD/NDD genes using in vivo and in vitro activity-by-contact (ABC) score analysis. We characterized the regulatory activity of these 5,425 candidate enhancers using MPRA, finding 3,793 functional enhancers. Motif analysis of functional enhancers identified transcriptional drivers of activity, and top active enhancers were prioritized as targets for CRISPRa. We next designed a library containing 9,688 sgRNAs targeting these 2,422 MPRA-prioritized enhancers, 3,044 sgRNAs targeting 761 isoform-specific ASD/NDD gene promoters, 1,512 additional promoter-targeting sgRNAs, and 1,500 non-targeting controls for a multiplex CRISPRa screen in iPSC-derived excitatory neurons.

(continue to page 56)

(continued from page 55)

This screen identified 512 highly effective targeting sgRNAs, which collectively upregulate 60% (203/337) of haploinsufficient ASD/NDD genes; this includes 100 enhancer-targeting sgRNAs which define 91 novel enhancer-gene pairs. In summary, we generated a catalog of experimentally validated, target-linked enhancers active in differentiating neurons, and discovered sgRNAs that can be used for the development of cis-regulation therapies for haploinsufficient forms of ASD and NDD.

Long read sequencing for enabling rare disease diagnosis

POSTER 303

Sarah L. Stenton*[1-2], Yulia Mostovoy*[1,3-4], Alba Sanchis-Juan*[1,3], Shloka Negi[5], Melanie O'Leary[1], Ben Weisburd[1], Elizabeth Carbonell[1], Grace VanNoy[1], Lalai Shah[1], Brian Mangilog[1], Eva Martinez1, Jillian Serrano[1], Gabrielle Lemire[1-2], Yongqing Huang6, William J. Rowell[7], Christopher T. Saunders[7], Emma Walker8, Stacey Hall[1], Kiran Garimella[6], Sheila Dodge[8], Stacey Gabriel8, Niall Lennon[8], Heidi L. Rehm[1,3], Harrison Brand[1,3-4,9], Karen H. Miga[5], Benedict Paten[5], Mike Talkowski[1,3-4,10], Anne O'Donnell-Luria[1-3]

1 Center for Mendelian Genomics, Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA
2 Division of Genetics and Genomics, Boston Children's Hospital, Harvard Medical School, Boston, MA
3 Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA
4 Department of Neurology, Massachusetts General Hospital, Boston, MA
5 Genomics Institute, University of California, Santa Cruz, CA
6 Data Sciences Platform, Broad Institute of MIT and Harvard, Cambridge, MA
7 Pacific Biosciences, Menlo Park, CA
8 Genomics Platform, Broad Institute of MIT and Harvard
9 Pediatric Surgical Research Laboratories, Department of Surgery, Massachusetts General Hospital, Boston, MA
10 Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA

More than 50% of families with suspected rare monogenic diseases remain unsolved after comprehensive genome analysis by short read sequencing (SRS). By capturing variants inaccessible to SRS, as well as methylation and variant phase by haplotype assembly, long read sequencing (LRS) is emerging as a technology with transformative potential to reduce the diagnostic gap. The potential added diagnostic yield of LRS over SRS remains unknown. Here, for >100 families enrolled in the Broad Institute Center for Mendelian Genomics Rare Genomes Project, we applied Oxford Nanopore (20 families, 61 individuals) or PacBio HiFi (83 families, 200 individuals) LRS.

(continue to page 58)

(continued from page 57)

These families - with neurodevelopmental, vascular, or musculoskeletal phenotypes - remained unsolved after SRS analysis. Both LRS technologies had median sequencing depth >30x, read length >15kb, and a modest increase in callable bases over SRS. A comprehensive array of methods - capturing single nucleotide variants (SNVs), small insertions/deletions (indels), structural variants (SVs), tandem repeat expansions, and methylation outliers - were applied, with analysis focused on rare variants using SRS and LRS reference population frequencies (gnomAD, All of Us, and/or the 1000 Genomes Project). A median of 69 rare non-synonymous coding LRS-only SNV/indels were detected per sample, 6 per sample within disease-associated genes, including a diagnostic de novo stop-gain SNV in CASK for a proband with neurodevelopmental delay. Optimizing for sensitivity, a median of 22k (PacBio) and 29k (Oxford Nanopore) SVs were captured per sample by LRS, approximately a two-fold increase over SRS. Among rare gene-disrupting LRS-only SVs was a diagnostic intragenic tandem duplication spanning multiple exons of ENG in a proband with hereditary hemorrhagic telangiectasia. In 7 families, previously overlooked diagnostic variants were detected by both SRS and LRS, for which LRS uniquely defined the breakpoints of multiple SVs and phased compound heterozygous and de novo variants, including the discovery of a highly recurrent de novo single base insertion in the novel non-coding disease gene, RNU4-2, found exclusively on the maternal allele. Work is ongoing to sequence additional families and complete analyses fully leveraging the unique capabilities of LRS. Our initial findings demonstrate the added utility of LRS for rare disease diagnosis.

Machine learning-driven liquid biopsy analysis of uterine lavage protein signatures that detects early stage ovarian and endometrial cancers

POSTER 304

Boris Reva[1], Dmitry Rykunov[1], Maria Padron[1], Deep Pandya[2], Swierczek Sabina[3], Linus Chuang[3], David Doo[3], Steve Sieber[3], Paul Fiedler[3], Caitlin Carr[1], Valentin Kolev[1], Konstantin Zakashansky[1], Vaagn Andikyan[4], Anton Iliuk[5], Peter Dottino[4,6], John A. Martignetti[1,6]

1 Icahn School of Medicine at Mount Sinai, New York, NY
2 Nuvance Health/Ruggles Institute, Poughkeepsie, NY
3 Nuvance Health, Poughkeepsie, NY
4 Yale University School of Medicine, New Haven, CT
5 Tymora Analytical Operations, West Lafayette, IN
6 MDDx, Inc., New York, NY

There is a clear unmet need for the development of screening and diagnostic tests to diagnose and detect early stage ovarian (OvCA) and endometrial (EndoCA) cancers prior to symptom onset and disease spread. This is that clinically pivotal timepoint when these cancers are more likely curable and not just treatable with poor 5-year survival rates. We have been developing a novel, machine learning driven, liquid biopsy-based screening test for diagnosing these cancers for the 82M+ perimenopausal and older women at risk and the 80,000 women who will be diagnosed with one of these cancers in the US annually.

To this end, we have developed a machine learning (ML)-based method to build a multiple protein biomarker classifier which allows diagnosis starting from a sample of uterine lavage fluid. Extracellular vesicles (EVs) from the fluid are isolated and their proteins analyzed by mass spectrometry (MS). Unlike traditional ML approaches that use gene expression levels as biomarkers, we exploited differences in the relative positioning of proteins in cancer and non-cancer samples and moved to dimensionless entropy scores. We enrolled 1,100 women from multiple institutions. Using our pipeline, the abundance levels of 1,170 proteins were quantified by MS. All AUCs for all tests sets – cancer vs non-cancer, OvCA v non-cancer, EndoCA v non-cancer and OvCA v EndoCA all exceeded 0.95 and sensitivities and specificities for the tests exceeded 90%. As verification, we tested our EndoCA classification scores on a completely independent data set using the NIH-funded, CPTAC data of EndoCA tumor tissues and normal controls. We obtained perfect differentiation between cancer and non-cancer samples.

(continue to page 60)

(continued from page 59)

We believe these findings set the stage for a transformational shift in women's health. Not only does our diagnostic platform support the possibility of early detection of OvCA and EndoCA with high sensitivity and specificity, with the shift toward value-based care models, reductions in expensive invasive radiologic and OR-based diagnostic approaches will be essential to drive down health care costs while maximizing value. Most importantly, we believe this platform provides the ability to detect early stage disease, thus the opportunity to make these cancers curable, not just treatable.

Models for characterization, diagnosis, and treatment of human cancers using comparative canine-human multi-omics

POSTER 305

Gregory Tawa[1], Gavrila Wicaksono[1], John Braisted[1], David Gerhold[1], Gurmit Grewal[1], Gurusingham Sittampalam[1], Christina Mazcko[1,4,5], Amy Leblanc[1,4,5], Douglas Thamm[6], Matthew Breen[7]

1 National Institutes of Health (NIH), Rockville, MD
2 National Center for Advancing Translational Sciences (NCATS), Rockville, MD
3 TDB, Rockville, MD
4 National Cancer Institute (NCI), Bethesda, MD
5 National Cancer Institute (NCI), Comparative Oncology Program (COP), Bethesda, MD
6 Colorado State University, Fort Collins, CO
7 North Carolina State University, Raleigh, NC

Abstract:
Most cancer modeling experiments are performed on rodents. While rodent models have been invaluable for investigating cancer mechanisms, they have not always been representative of the disease in humans. The use of companion animals to understand human tumor biology stems from decades of scientific observation that pet dogs spontaneously develop malignancies that share characteristics with human cancers. In this context, we developed an experimental/bioinformatics pipeline that generates and analyzes multi-disease, multi-species, next-generation sequencing data, identifies relevant disease genes common to the species analyzed and based on these common genes enumerates potential drug targets and therapeutic drug combinations. Validation of candidate genes is conducted using proteomics experiments to ensure that they express as proteins in tissue and existing drugs are identified that are known to modulate protein expression opposite to that exhibited in the disease. Drug combinations are derived such that each drug modulates different proteins to maximize synergy. Matrix screens are performed on canine cell lines and combinations that show significant efficacy and synergy are selected for further study in mouse PDX models with the eventual goal of translating these combinations into in vivo canine and human studies. Using this strategy, we found 68 protein targets and 6 drug combinations exhibiting synergy for canine osteosarcoma. Mouse PDX experiments at NCI are ongoing to evaluate these six drug combinations in vivo. This work exemplifies an approach that further establishes the relevance of canine to human cancer and provide opportunities to explore cancer mechanisms and treatments in both species.

Molecular optimization of a cell free DNA enrichment assay for deep exome sequencing

POSTER 306

John Lemanski[1], Christopher W. Whelan[1,3,4], Michael Duyzend[1,3,6], Monica Salani[1], Stephanie P. Hao[1,5], Elise Valkanas[1,3], Caroline Cusick[4], Krithika Nathamuni[1], Harrison Brand[1,2,3,4,5], Mike Talkowski[1,2,3,4]

1 Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA
2 Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA
3 Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA
4 Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA
5 Pediatric Surgical Research Laboratory, Massachusetts General Hospital, Boston, MA
6 Department of Pediatrics, Boston Children's Hospital, Boston, MA

The assessment of cell free DNA (cfDNA) extracted from maternal plasma has revolutionized prenatal diagnostics. We recently developed a high resolution non-invasive fetal sequencing (NIFS) approach capable of capturing the fetal exome from a maternal blood draw alone and applied this technique at all stages of pregnancy (Brand et al. 2023, NEJM). Here, we highlight molecular innovations and optimizations to NIFS that greatly increase scalability, reliability, and efficiency. Traditionally, non-invasive testing during pregnancy has been hampered by the low variant allele frequency of mutated DNA, necessitating samples undergo sequencing at a high depth to achieve sufficient coverage for accurate genotyping. Additionally, the scarcity of cfDNA serves as a significant bottleneck to achieving high library complexity, which in turn limits the coverage obtainable, as the duplication rate of the amplified libraries reduces the rate of discovery as sequencing depth increases. To overcome these challenges, we performed a molecular optimization which sought to maximize the conversion of maternal plasma into sequenced molecules. We tested several column and bead based cfDNA extraction kits (Qiagen, Perkin Elmer and Revvity) to maximize the recovery of cfDNA fragments finding the QIAamp Min elute ccfDNA most effective. Furthermore, to minimize the rate of molecule decay across the library workflow we increased the molar ratio of adapter to DNA insert 10 fold from standard methods to account for the shorter fragment size of cfDNA, compared to sheared genomic DNA common with dA-tailing protocols, and found a dosage sensitive 52-56% change in library yield across fixed inputs with no adapter dimer formation. Finally, we extended the length of the hybridization of our target enrichment from 4 hours to an overnight incubation which relatively reduced off target rate by 18% and increased yield of captured molecules by 109%. These modifications lead to a 267% increase in library complexity from our previously established method, which subsequently, allowed for a 1.7 fold increase in relative sequencing depth due to a 43% relative reduction in the duplication rate. Overall, this work highlights several key pain points in the NIFS method and demonstrates the importance of library optimization for maximizing NIFS throughput.

New insights for prenatal genomics: deep sequencing reveals extensive spatial and temporal genetic heterogeneity of placenta

POSTER 307

Ieva Miceikaite[1,3], Christina Fagerberg[2], Charlotte Brasch-Andersen[1,2], Qin Hao[2], Pernille Mathiesen Tørring[2], Britta Schlott Kristiansen[2], Martin Jakob Larsen[1,2]

1 Clinical Genome Center & Human Genetics, Department of Clinical Research, Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark
2 Department of Clinical Genetics, Odense University Hospital, Odense, Denmark
3 Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA

Confined placental mosaicism (CPM) is a well-known phenomenon that can influence prenatal genetic testing outcomes. Characterized by genetic discrepancies between placental and fetal cells, CPM occurs in 1-2% of pregnancies for chromosomal aberrations, with recent studies indicating sequence-level changes as well. This study aimed to explore the genetic heterogeneity of placentas at different locations and stages of pregnancy using deep genome sequencing.
We compared whole genome sequencing results from first-trimester chorionic villi (CVS) or second-trimester amniotic fluid (AF) samples, four placental biopsies, and a fetal tissue sample from each of the six cases in the study. In five cases, deep exome sequencing was performed on circulating DNA from maternal plasma.
Analysis revealed that most de novo sequencing variants in the fetus were present in all placental samples. Distinct de novo variant signatures among placental biopsies indicated subclonal evolution and chronological changes. None of the subclonal de novo variants were found in the fetus, and maternal contamination was identified in one placental biopsy.
Our study provides new insights into the genetic heterogeneity of the placenta, highlighting the presence of de novo sequence changes in both coding and non-coding regions, and emphasizing the extent of chronological changes. These findings enhance our understanding of sequence-level mosaics in prenatal diagnostics, underscoring the need for careful interpretation of placental genetic data to avoid misdiagnoses.

Noncoding gene discovery for neurodevelopmental disorders

Abstract
Selected
Talk

Anne O'Donnell-Luria[1,2]

1 Broad Institute of MIT and Harvard, Cambridge, MA
2 Boston Children's Hospital, Harvard Medical School, Boston, MA

Noncoding RNAs comprise a large fraction of the human genome, yet few have been shown to be associated with human disease. Two recent noncoding gene discoveries highlight the importance of evaluation of the noncoding genome.
Haploinsufficiency due to Alu-Alu mediated de novo deletion of the long noncoding RNA CHASERR was identified in 3 children with severe developmental delay, facial dysmorphisms, and cerebral hypomyelination (Ganesh et al., medRxiv, 2024). CHASERR has previously been shown to be important for the negative regulation of CHD2, a haploinsufficient disease gene associated with developmental and epileptic encephalopathy. We showed increased CHD2 protein expression in these children, demonstrating bidirectional dosage sensitivity of CHD2.
De novo variants in an 18 bp critical region of RNU4-2, which encodes the U4 small nuclear RNA (snRNA) component of the major spliceosome, have been identified in 115 individuals with a severe neurodevelopmental phenotype with distinct facial features and multiorgan involvement (Chen et al., medRxiv, 2024). Most (77%) have the same highly recurrent single base-pair insertion (n.64_65insT). Individuals with SNVs appear to be more mildly affected than those with insertions.
By RNA-sequencing analysis, we found systematic disruption of 5' splice site usage in individuals with RNU4-2 pathogenic variants. This adds to the signatures we can identify for spliceosome disorders, including a pattern of minor intron inclusion in the syndromic disorder of the minor spliceosome due to biallelic pathogenic variants in RNU4ATAC. We hypothesize that these signatures can be expanded to improve diagnosis of established spliceosome conditions and towards identifying additional novel conditions in noncoding RNA not yet associated with human disease.
The majority of patients with rare disease remain undiagnosed and disruption of noncoding genes likely accounts for a sizeable fraction of etiologies, highlighting the importance of building large broadly consented genomic datasets of undiagnosed rare disease patients to empower continued discovery.

Non-linear machine learning models to improve variant prediction among clinical breast cancer patients

POSTER 308

Larah Siouffey[1], Adam Chamberlin[2], Amal Yusuf[3], Marcy Richardson[3], Linda M. Polfus[2]

1 Keck Graduate Institute, Department of Human Genetics and Genomic Data Analytics, Claremont, CA
2 Ambry Genetics, Advanced Analytics, Aliso Viejo, CA
3 Ambry Genetics, Medical Sciences, Aliso Viejo, CA

Approximately 13% of women will be diagnosed with breast cancer at some point during their lifetime. Women harboring pathogenic variants (PV) in key homologous repair deficiency genes have both increased breast cancer risk and respond more favorably to certain therapies, like platinum agents and PARP inhibitors. Traditional breast cancer risk models are enhanced by Polygenic Risk Scores (PRS), yet often fail to integrate multiethnic data and complex SNP interactions. We examined (N=10,880) ethnically diverse patients who underwent a multi-gene panel test for cancer predisposition in a single clinical diagnostic laboratory. Patients were selected for having a rare PV among one or more of five breast cancer predisposition genes (ATM, BRCA1, BRCA2, PALB2, and CHEK2) as well as 860 common SNPs known to be associated with breast and other cancers. Cases were breast-cancer-affected according to test requisition forms and clinical notes while controls unaffected by breast cancer and reported no family history of breast cancer. Diverse ethnic groups included Caucasian (67.2%), Ashkenazi Jewish (5.8%), African American (10.5%), Hispanic (11.1%), and Asian (5.5%). We applied LASSO regression to identify influential features based on age at diagnosis (binned by decade), self-reported ethnicity, PV status, and SNP data, achieving an AUC of 0.81 (Mean Square Error: 0.18; r2: 0.26). The LASSO model's significant predictors (N=139) were refined through XGBoost, which utilized an 80/20 training/testing split, resulting in a model AUC of 0.83. Predictions on the test set (N=2176) yielded an AUC: 0.83 (Percent Variance Explained: 28.6%; r2: 0.40). XGBoost selected 134 features and ranked importance revealed 5 PV genes, age group, Caucasian, Ashkenazi Jewish ancestry, rs1329390, rs11814448, rs2660753, and rs6465657. The XGBoost model underscored several intergenic SNPs and others co-localizing with known breast cancer-associated regions. A compelling interaction was detected between CHEK2 PVs and rs132390, an intronic variant within EMD1, a gene implicated in cancer metastasis. Future research will expand the dataset via ICD10 phecodes and incorporate additional machine learning features into an enhanced PRS framework, integrating the Tyrer-Cuzick model to refine risk predictions. Our approach aims to accurately identify individuals at increased risk, optimizing preventive and therapeutic interventions based on personalized genetic profiles.

Oncogenicity classification of 100 pediatric cancer variants using a standardized assessment framework: lessons learned

POSTER 309

Wesley A. Goar[1], Kori Kuzma[1], Kathryn Stahl[1], Michael D. McCarrick[1], Kathleen M. Schieffer[1,2], Catherine E. Cottrell[1,2], Elaine R. Mardis[1], Alex H. Wagner[1,3]

1 The Steve and Cindy Rasmussen Institute for Genomic Medicine at Nationwide Children's Hospital, Columbus, OH
2 Departments of Pathology and Pediatrics, The Ohio State University, Columbus, OH
3 Departments of Pediatrics and Biomedical Informatics, The Ohio State University, Columbus, OH

The ClinGen/CGC/VICC Oncogenicity guidelines were published as a standardized approach for assessing genomic variation and its relationship to the promotion of cancer formation. We describe the Variation Categorizer (VarCat), an assessment tool designed to meet the rigor of clinical grade variant classification that creates highly-structured oncogenicity classifications following these guidelines through an intuitive user-facing web application. VarCat was designed in collaboration with our clinical team at Nationwide Children's Hospital to enable precise and reproducible variant classifications for use in clinical workflows. The resulting variant classification is stored in a standardized genomic knowledge format developed by the Global Alliance for Genomics and Health (GA4GH) to promote interoperability and data sharing across the pediatric cancer community. A demo version of this tool will be available for trial use and feedback. We plan to make this tool open-source to promote data sharing among pediatric institutions worldwide.
We report what was learned during our assessment of 100 somatic variants from pediatric/AYA cancer cases studied by NGS-based exome sequencing in relation to prior assessments of these variants as characterized based upon the AMP/ASCO/CAP clinical actionability guidelines. Our study assesses application of these complementary guidelines in a high-throughput clinical setting and the advantage in joint utilization to perform variant classification. We describe the frequency and impact of specific oncogenicity codes used in these variant classification assessments and examine our final oncogenicity classification in relation to the previously assessed clinical actionability tiering classification. Our findings highlight specific oncogenicity codes that would benefit from further clarification for consistent application within clinical classification workflows. We discuss the need for objective and explicit criteria that classify cited studies as providing well-established, reproducible, and robust evidence. These criteria enable modulation of the strength of functional evidence, allowing for weaker evidence to be applied objectively and consistently. We present the challenges created due to these subjective guidelines and present our internal framework for systematically modifying and applying these codes in our clinical variant classification workflows.

Optics-free spatial mapping of tissues through indexed sequencing

POSTER 310

Abdulraouf Abfulraouf[1], Weirong Jiang[1], Wei Zhou[1], Junyue Cao[1]

1 The Rockefeller University, New York, NY

The intricate structure and varied composition of cells in various organs and tissues present challenges in systematically characterizing intricate molecular and cellular interaction mechanisms of diseases and aging. Even with the recent rise of spatial genomic techniques, mapping the spatial locations of cell types and their interaction network remains difficult due to cost and throughput limitations. Here, we have developed a novel methodology called IRIS (Imaging Reconstruction using Indexed Sequencing) that enables cost-effective spatial transcriptomics profiling without relying on optical imaging. Through DNA microscopy-based neighborhood interaction tissue reconstruction, IRIS allows extensive analysis of large tissue sections and many replicates with adjustable mapping resolution at only a fraction of the cost of other commercial platforms. With the IRIS platform, we reconstructed a large 2D spatial array with two whole mouse brain coronal sections (2.5+ cm^2). Moreover, we also created a spatially resolved transcriptome atlas of the mouse brain and identified aging-associated changes in gene expression and spatial organization across various brain regions. Further analysis of cell-cell interaction changes identified aging-associated foci in white matter regions enriched with inflammatory subtypes of microglia and oligodendrocytes. Overall, the IRIS methodology cost-effective and ease-of-use approach makes it broadly applicable to the studies of spatial gene expression changes in various systems.

Optimized DNA library preparation workflows to overcome challenges common to critical sample types

POSTER 311

Dustin Hite[1], Giulia Corbet[1], Philip Benson[1], Josh Haimes[1], Lindsay Peterkin[1], Martin Ranik[1], Andrew Gaines[1], Thomas Harrison[1], Kailee Reed[1], Kristina Giorda[1], Brian Kudlow[1]

1 Watchmaker Genomics, Boulder, CO

NGS has been effectively used in clinical oncology and translational research for various diagnostic purposes. These include identifying hereditary conditions, monitoring disease progression, and guiding personalized treatment approaches, such as minimal residual disease (MRD) monitoring. To overcome challenges related to clinical sample quality and quantity, high-performance enzymes and streamlined workflows are essential for these assays. By utilizing highly optimized library preparation methods, more molecules can be sequenced with less bias, ultimately increasing sensitivity and specificity for a wide range of applications.

Various DNA library prep methods were assessed for their effectiveness in working with clinically relevant sample types and applications. The matched inputs varied across kits, ranging from 75 to 300 ng gDNA, 50 to 200 ng of FFPE DNA, and 1 to 25 ng of mock cfDNA – input masses typical in clinical applications. Watchmaker solutions across sample input types and input masses produced high yield along with low indels, softclips, chimeras, improper pairs, and hairpin artifacts. PCR-free whole genome sequencing resulted in highly accurate variant calling accuracy F1 scores. In addition, Watchmaker libraries showed the highest coverage in hard-to-cover promoter regions (Ross et al., 2013). A standardized method for library preparation from FFPE samples was developed to produce similar insert sizes across varying input mass and quality of FFPE samples. Enzymatically fragmented samples virtually eliminated hairpin artifacts that were present in up to 4.5% of reads in sonication-based libraries. Furthermore, the mean target coverage with enzymatic fragmentation is significantly higher relative to sonication, when normalized to pre-sheared DNA input. Finally, target coverage with mock cfDNA libraries prepared with the Watchmaker DNA Library Prep Kit was higher than competitor kits across all input ranges tested. Variants were accurately detected by UMI-based analysis even at the lowest expected variant allele fraction.

These workflows enable high-quality DNA library preparation from a range of sample types and inputs, producing high coverage, uniform insert sizes, and minimal sequencing artifacts which improves sensitivity and specificity. This approach is highly scalable and automatable, enabling various oncology applications.

Optimizing recall of RNA-seq outliers in heterogeneous rare disease cohorts

POSTER 312

Joshua L. Major-Mincer[1,3], Aaron R. Quinlan[1-3]

1 University of Utah, Department of Biomedical Informatics, Salt Lake City, UT
2 University of Utah, Department of Human Genetics, Salt Lake City, UT
3 University of Utah, Utah Center for Genomic Discovery, Salt Lake City, UT

Rare diseases collectively affect millions of individuals worldwide, yet half of cases remain unresolved. This diagnostic gap is partially due to the prevalence of variants of uncertain significance (VUS), which lack sufficient prior evidence connecting them to a disease phenotype. As a result, computational tools may inadequately predict the consequences of novel VUS.

In principle, RNA sequencing (RNA-seq) can address these computational shortcomings by providing a direct transcriptional readout of a VUS's impact on total transcript and isoform abundance. In rare disease studies, an important challenge is identifying a patient's expression as an "outlier" with respect to a reference RNA-seq population. A prominent challenge to this objective is that rare disease populations are often small yet heterogeneous in age, sex, phenotype, and technical processing. Such heterogeneity attenuates the power to detect true outliers in expression or splicing.

Recognizing these challenges, computational approaches have been used or developed to control noise in rare disease expression data, including existing logarithmic transformations or the innovative OUTRIDER autoencoder normalization framework. Nonetheless, we argue that heterogeneous populations and inadequate sample sizes remain persistent challenges. Preliminary analyses with the Genotype-Tissue Expression (GTEx) dataset indicate a weaker correlation in sample outlier statistics when comparing bootstrapped reference populations consisting of both males and females against those consisting of a single, uniform sex. We then hypothesize that biological and technical confounders result in greater false negative expression outlier calls even after correction with existing approaches. To test this hypothesis, we use whole blood RNA-seq expression data from the GTEx dataset and simulate outliers at various magnitudes of deviance from the raw population mean. Specifically, outliers are injected in genes associated with changes in expression due to phenotypic traits in healthy populations. Groups of samples are bootstrapped at several population sizes, and outliers in expression are identified for each group using the described methods. Finally, the change in recall of outliers is measured as a function of phenotypic confounders and reference population size. We anticipate our findings will empower confident utilization of RNA-seq in rare disease diagnostics through reasoned and considerate cohort structure.

Personalized prediction of an autistic child's probabilities of carrying a clinically diagnosable variant

POSTER 401

Susan Kuo[1,2], Celia van der Merwe[1,2], Jack Fu[1,2], F. Kyle Satterstrom[1,2], Mike Talkowski[1,2,3], Audrey Thurm[4], Stephan Sanders[5,6], Somer Bishop[6], Elise Robinson[1,2,7,8]

1 Broad Institute of MIT and Harvard, Stanley Center for Psychiatric Genetics, Cambridge, MA
2 Massachusetts General Hospital, Harvard Medical School, Center for Genomic Medicine, Boston, MA
3 Massachusetts General Hospital, Harvard Medical School, Department of Neurology, Boston, MA
4 National Institute of Mental Health, Neurodevelopmental and Behavioral Phenotyping Service, Bethesda, MD
5 University of Oxford, Department of Paediatrics, Oxford, England, United Kingdom
6 University of California San Francisco, Department of Psychiatry, San Francisco, CA
7 Harvard T.H. Chan School of Public Health, Department of Epidemiology, Boston, MA
8 Massachusetts General Hospital, Harvard Medical School, Department of Psychiatry, Boston, MA

US professional medical guidelines recommend clinical genetic testing for all children with autism diagnoses, yet only 3 in 100 autistic children have undergone first-line testing. Although a genetic diagnosis can lead to tailored services for an autistic child, families and clinicians often lack access to reliable, accurate sources of information about clinical genetic testing. To empower learning about genetics tailored to autism, we are building a public, web-based suite of educational resources named MINERVA. In consultation with expert advisory groups and focus groups of parents and pediatricians from the autism community, MINERVA will provide several empirically supported resources centered on genetic heterogeneity and clinical genetic testing. These resources include a novel prediction algorithm to estimate an autistic child's personalized probabilities of carrying a neurodevelopmental disorder associated variant (NDAV). NDAVs were defined as de novo copy number deletions, as well as de novo protein truncating variants and disruptive missense variants in genes associated with neurodevelopmental disorders (including intellectual disability and autism). To develop the prediction algorithm, we harmonized whole exome sequencing and phenotypic data in a combined sample of 3,440 autistic children aged 4-17 years from two independent, North American cohorts.

(continue to page 71)

(continued from page 70)

We conducted logistic regression models using multiple, easily accessible phenotypes (e.g., sex, age at walking) to estimate genetic diagnosis probabilities. We observed extensive heterogeneity in probability estimates throughout our sample, ranging from approximately 2% in males without congenital heart defect who walked at 12 months and had a cognitive score above 110, to over 50% in females with congenital heart defect who started walking at 30 months and had a cognitive score below 70. MINERVA will present genetic diagnosis probabilities for a child along with average probabilities across autistic people (7%) and their undiagnosed siblings (1%). By providing these probability estimates along with accessible guides to current scientific and clinical knowledge about autism genetics, MINERVA can help catalyze informed decision-making around clinical genetic testing for autistic children.

Population screening for actionable genomic conditions and pharmacogenomics in primary care: the Alabama Genomic Health Initiative

Abstract
Selected
Talk

Nita A. Limdi[1], Blake Goff[1], Brittney Davis[1], Irene Moss[1], Irfan Asif

1 Heersink School of Medicine, University of Alabama at Birmingham, Birmingham, AL

Introduction: Supported by the State of Alabama, the Alabama Genomic Health Initiative (AGHI) provides free genomic testing, interpretation, and counseling for disease risk and pharmacogenetics for all Alabamians.

Methods: AGHI (2017-2020) enrolled Alabamians from all 67 counties and conducted population screening for actionable genomic conditions (ACMG SF v2.0), recommendations for patients harboring genomic risk. In 2021, testing was expanded to include genomic screening (ACMG SF v3.1) and pharmacogenetics (PGx; 26 genes in 97 markers; meeting FDA/CPIC evidence thresholds) and deployed in 3 primary care settings.

A multi-disciplinary team of physicians, genetic counselors and pharmacists was operationalized to return results to providers and patients. A Genomic Community Advisory Board engaged diverse perspectives, experiences, and communities to guide an implementation science informed evaluation of barriers and facilitators in genomic medicine integration

Results: We engaged over 8000 participants, returning 115 actionable genomic disease risk reports (since 2017) and 1350 PGx reports (since 2022).

About 2% patients harbor disease risk variants associated with increased predisposition to cardiac diseases/hypercholesterolemia (n=59), cancers (n=40), and other (n=16) conditions. Among findings, 36% (N=41) were associated with increased risk for a CDC recognized Tier 1 condition, including hereditary breast and ovarian cancer (19%), Lynch syndrome (6%), Hereditary TTR amyloidosis (9%) and familial hypercholesterolemia (10%).

(continue to page 73)

(continued from page 72)

Pharmacogenetic actionability: All (100%) participants harbor an actionable PGx variant. Based on pharmacist conducted medication reconciliation, 70% were on medication with PGx guidance, 48% harbor PGx variants and were taking an affected medication (e.g. antidepressants, NSAIDS, proton-pump inhibitors and tramadol). In 10% of participants, pharmacists sent an active alert to the provider to recommend alternative medication.

Evaluation of barriers and facilitators: Acceptability of genomic medicine and alignment with the clinic's values was strongly viewed but not about its feasibility. Thematic analysis of semi-structured interviews identified 3 key challenges: tension between providing more information and time constraints; lack of technology to integrate information into workflow; and uncertainty about point-of-care actionability of results. Education about genomics and highlighting its value for prevention can facilitate overcoming these challenges.

Conclusions: We share our experience on population screening for common actionable genomic conditions and pharmacogenetics in primary care and highlight the broader implications for the implementation of genomic medicine in primary care in the United States.

Abstract
Selected
Talk

Preemptive return of clinical pharmacogenomic results to over 50,000 individuals in a population-scale biobank: Phenotype and actionable medication exposure characteristics

James L. Martin[1,2], Kristy R. Crooks[1,3], Casey S. Greene[1,3], Ashley Hansen[4], Emily C. Hearst[1,4], Kaitlyn W. Hess, MS, MLS(ASCP)[4], Natalie Johnson[4], David P. Kao[1,3], Elizabeth L. Kudron[1,3], Yee Ming Lee[1,2], Nicole L. McDaniel[1,2], Nicholas Rafaels[1], Elise L. Shalowitz[1], Carolyn Swartz[4], Anna Tanaka[1,4], Katy Trinkley[1,3], Sharon Vandenberg[4], Christina L Aquilante[1,2]

1 Colorado Center for Personalized Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO
2 University of Colorado Skaggs School of Pharmacy and Pharmaceutical Sciences, Aurora, CO
3 University of Colorado, School of Medicine, Aurora, CO
4 University of Colorado, UCHealth, Aurora, CO

Introduction: Among population-scale biobanks, some offer the opportunity to return clinical pharmacogenomic (PGx) results to the electronic health record (EHR) for use in clinical care. Here, we describe PGx phenotype and actionable medication exposure characteristics of >50,000 participants who have received results from the biobank at the Colorado Center for Personalized Medicine (CCPM biobank).

Methods: The CCPM biobank preemptively returns clinical PGx results electronically to the EHR for CYP2C19, SLCO1B1, DPYD, CYP2C9, TPMT, NUDT15, and ABCG2. Automated drug-gene interaction (DGI) EHR alerts provide clinicians guidance when prescribing medications to participants with at-risk phenotypes. We evaluated CCPM biobank participant demographic, phenotype, and alert characteristics from 12/01/21-02/25/24.

Results: Over the 27-month period, 51,649 distinct participants had at least one PGx result returned to the EHR, totaling 325,469 distinct PGx results (Epic Genomic Indicators). The cohort was 61.6% female, 85.2% of European descent, and 9.4% Hispanic; mean ± SD age at the time of result return was 53.7 ± 16.7 years, and 99.2% of participants were alive as of 2024. As expected, the frequency of non-normal phenotypes was highest for CYP2C19 (58.8%), followed by CYP2C9 (34.6%), SLCO1B1 (27.1%), and ABCG2 (21.8%). There were 42,947 (83.2%) distinct participants with results returned for all seven genes.

(continue to page 75)

(continued from page 74)

The percentage of participants with ⩾1, ⩾2, or ⩾3 non-normal phenotypes for at least one of the seven genes was 90%, 50.6%, and 15.4%, respectively. Of the 51,649 participants with at least one PGx result returned, 5,794 (11.2%) had at least one distinct DGI alert, indicating a potentially actionable medication exposure. Of all distinct DGI alerts (n=7,247), the most common were omeprazole/CYP2C19 (26.5%), pantoprazole/CYP2C19 (21.1%), atorvastatin/SLCO1B1 (14.6%), rosuvastatin/SLCO1B1/ABCG2 (9.4%), and escitalopram/CYP2C19 (6.8%).

Conclusion: Our findings demonstrate that population biobanks can be successfully leveraged to deliver precision medicine at scale, with immediate clinical impact across a variety of drug classes.

Abstract
Selected
Talk

Prequalification of genomic newborn screening for severe childhood genetic diseases using harmonized population and variant databases

Meredith S. Wright[1,2], Yupu Liang[3], William R. Mowrey,[3] Mia Gover[1], Liana Protopsaltis,[1,2] Katarzyna Ellsworth,[1,2] Mei Baker,4 Annette Feigenbaum,[1,2,5] Erwin Frise,6 Lucia Guidugli,[1,2] Kevin P. Hall,[7] Christian Hansen,[1,2] Mark Kiel,[8] Lucita Van Der Kraan,[1,2] Hugh Kwon,[1,2] Jeremy Leipzig,[9] Lakshminarasimha Madhavrao,[1,2] Rebecca Mardach,[1,2,5] Danny Oh,[1,2] Lauren Olsen,[1,2] Rebecca Reimers[1,2], Jennifer Schleit[1,2], Shyamal S. Mehtalia,[7] Kristen Wigby,[1,2] Mary J. Willis,[1] Mark Yandell[10], Chris M. Kunard,[7] Thomas Defay,[3] Stephen F. Kingsmore[1,2]

1 Rady Children's Institute for Genomic Medicine, San Diego, CA
2 Rady Children's Hospital, San Diego, CA
3 Alexion, AstraZeneca Rare Disease, Boston, MA
4 Department of Pediatrics, University of Wisconsin School of Medicine and Public Health, Madison, WI
5 Department of Pediatrics, University of California San Diego, San Diego, CA
6 Fabric Genomics, Inc., Oakland, CA
7 Illumina, Inc., San Diego, CA 92122
8 Genomenon Inc., Ann Arbor, MI
9 TileDB Inc., Cambridge, MA
10 Department of Human Genetics, University of Utah, Salt Lake City, UT

Newborn genome sequencing can build upon the success of existing newborn screening (NBS) to improve outcomes for individuals with selected, severe childhood conditions by identification and treatment before symptom onset. We designed a platform (BeginNGS) to accomplish this based on NBS with a derivative of rapid diagnostic genome sequencing, artificial intelligence, and a management guidance system. To meet this objective, we tested BeginNGS, an automated variant identification and decision support system for effective implementation of population scale newborn genome sequencing. Harmonized variant sets from Clinvar, literature, and locus-specific variants databases resulted in over 54,000 candidate pathogenic (P) and likely pathogenic (LP) variants associated with 409 disorders and 342 genes. Here we report results of a retrospective study of the sensitivity and specificity of the platform across a diverse population of 500,000 UK Biobank (UKBB) and 150,000 Mexico City Prospective Study (MCPS) subjects, and 7,575 individuals who had received RDGS in our clinical diagnostic laboratory. Training with BeginNGS identified variants and haplotypes with false positive P and LP assertions which were prequalified as non-causal for single locus genetic disorders due to frequencies inconsistent with disease prevalence, penetrance, expressivity, locus heterogeneity, and inheritance mode.

(continue to page 77)

(continued from page 76)

Retrospective review indicated that BeginNGS had a sensitivity of 97% for individuals who received rapid diagnostic genome sequencing with P or LP variants in BeginNGS genes using a combination of pre-qualitifed P and LP variants and a modified AI model for variant identification. A comparative analysis of screen positivity rate when expanding gene-condition dyads to the full set of 1354 different genes proposed by seven different global newborn genome sequencing initiatives versus restricting to the core 128 genes shared amongst the research studies is predicted to reveal a range in projected positivity rates that have implications for clinical utility and outcome measures. The BeginNGS studies suggest a generalizable approach to validation of pathogenicity assertions associated with rare and ultra-rare genetic disorders. The data indicate that our platform is sufficiently mature to proceed to a large, adaptive clinical trial to evaluate clinical utility, acceptability, and cost effectiveness of newborn sequencing as a supplement to standard state NBS.

Prioritization and functional characterization of non-coding regulatory variants in congenital heart disease

POSTER 402

Neil Slaven[1], Sarah Morton[2], Yuri Kim[2], Michal Kosicki[1], Momoe Kato[1], Ingrid Plajzer-Frick[1], Christine Seidman2, Jonathan Seidman[2], Len Pennacchio[1], Axel Visel[1]

1 Lawrence Berkeley National Lab, Environmental Genomics and Systems Biology, Berkeley, CA
2 Harvard Medical School, Department of Genetics, Boston, MA

Congenital heart disease (CHD) encompasses a heterogeneous collection of structural cardiovascular defects, making up the most common class of birth defects impacting millions of individuals in the US alone. Despite efforts to identify the genetic determinants of the disease through large scale sequencing projects it is estimated that coding variants only account for ~30% of cases. As such, the Bench to Bassinet (B2B) and Pediatric Cardiac Genomics Consortium (PCGC) in partnership with TopMed have employed whole genome sequencing on large cohorts of patients. This effort has identified and sequenced over 1,831 trios in which the child has severe cardiac defects and no evidence for coding mutations, identifying over 75,000 de novo variants (DNVs) and many fold more rare variants in the non-coding genome. Previous studies showed that non-coding variants in regulatory elements, such as transcriptional enhancers, can in principle play a causal role in congenital diseases. However, the global contribution of variants occurring in heart enhancers as a harbor of CHD pathogenicity remains unclear. To address this knowledge gap, we aimed to prioritize candidate variants in predicted heart enhancers and characterize resulting changes to their function in vivo. We intersected DNVs and rare variants with an integrated catalog of over 100 bulk tissue H3K27Ac ChIP-Seq, DHS-Seq, ATAC-Seq, and single-cell ATAC-Seq datasets derived from human and mouse heart. This intersection identified ~1,000 DNVs and ~120,000 rare variants occurring in predicted heart enhancers. Variants were further prioritized based on their predicted impact on transcription factor binding site affinity obtained using the MotifbreakR software. In vivo testing of wildtype and variant enhancer sequences was conducted using a safe harbor reporter knock-in mouse assay (enSERT). To date our characterization has validated 14 developmental heart enhancers and identified 4 variants that reduce enhancer activity. These include 2 variants occurring within the same intronic enhancer of the TTN gene. To characterize the contribution of these, and future, variants to the development of heart defects we will conduct variant enhancer knock-ins followed by in-depth phenotyping assays.

RaftSeq distinguishes pathogenic from benign variants in the IDS gene

POSTER 403

Patricia Dickson[1], Anushka Viswanathan[1], Serena Elia[1], Steven Le[1], Balraj Doray[1], Sarah Hurt[1], William Buchser[1]

Washington University School of Medicine, St. Louis, MO

Pathogenic loss-of-function variants in the IDS gene cause the X-linked disorder, mucopolysaccharidosis type II, also known as Hunter syndrome. IDS encodes the soluble lysosomal enzyme, iduronate 2-sulfatase (I2S). Because only a few percent of normal I2S activity is thought to be corrective for Hunter syndrome, in silico models may overestimate the impact of potentially damaging variants. In addition, the typical in vitro assay uses an artificial substrate and is prone to identifying pseudodeficiency alleles. To improve VUS classification, we used RaftSeq, a functional genomics technology employing high-content imaging combined with machine learning to characterize and quantify cellular phenotypes. Human A549 cells were engineered to express known pathogenic (G224R and 228fsX) and known benign (V223I) IDS variants, grown in isogenic lines, and confirmed by sequencing and I2S enzymatic activity. Cells were plated on microrafts and imaged with fluorescent confocal microscope to obtain intracellular features of nuclei and lysosomes. Combining lysosomal measurements for intensity, puncta morphology and texture generated a 'pathogenic score.' Using this system, 70% of G224R variant cells, 68% of 228fsX variant cells, and 30% of V223I (benign) variant cells scored as pathogenic-like. Next, we produced and purified recombinant I2S and added it to the media in two concentrations to rescue the cellular phenotype. Successful uptake and correction of intracellular I2S was confirmed by intracellular assay. G224R and 228fsX variant cells showed phenotypic rescue, with 50 or 59% of cells scoring as pathogenic-like 144h after I2S treatment, respectively. Cellular phenotyping using RaftSeq may provide a solution to the problem of variant classification in lysosomal storage disorders such as Hunter syndrome.

Rare variant rates and its association with autism and neurodevelopmental phenotypes in Kenya and South Africa

POSTER 404

Patricia Kipkemoi[1,2,3], Mutaz Amin[4], Sarah L. Stenton[4,5], Emily O'Heir[4], Christina Austin-Tse[4,6], Ikeoluwa Osei-Owusu[4], Brigitte Melly[7], Emma Eastman[7], Zandre Brauer[7], Rizqa Sulaiman-Baradien7,8, Serini Murugasen[7], Claire Fourie[7], Sam Bryant[4], William Baddoo[4], Rachael Ivison[4], Martha Kombe[1], Alfred Ngombo[1], Constance Rehema[1], Moses Mangi[1], Collins Kipkoech[1], Pauline Mwatsuma[1], Eunice Chepkemoi[1], Alice Galvin[4], Celia van der Merwe[4], The NeuroDev Project1,[3,4,7], Charles Newton[1,3,9], Anne O'Donnell-Luria[4,5,6], Kirsten A Donald[7,10], Elise Robinson[4,6], Amina Abubakar[1,3,9]

1 Neuroscience Unit, KEMRI-Wellcome Trust, Center for Geographic Medicine Research Coast, Kilifi, Kenya
2 Complex Trait Genetics Department, Center for Neurogenomics and Cognitive Research, Vrije Universiteit, Amsterdam, Netherlands
3 Institute of Human Development, Aga Khan University, Nairobi, Kenya
4 The Broad Institute of MIT and Harvard, Cambridge, MA
5 Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA
6 Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA
7 Department of Paediatrics and Child Health, Red Cross War Memorial Children's Hospital and University of Cape Town, Rondebosch, South Africa
8 Departments of Pediatrics and Child Health, University of Manitoba, Winnipeg, Canada
9 Department of Psychiatry, University of Oxford, London, United Kingdom
10 Neuroscience Institute, University of Cape Town, Groote Schuur Hospital, Observatory, South Africa

Autism is a neurodevelopmental condition that has been reported to affect 1 in 100 children. Over 90% of children under the age of five years with neurodevelopmental disorders (NDDs), such as autism, live in lower and middle-income countries. However, the majority of autism research is carried out in high-income countries. We aim to investigate rare de novo variant burden and its impact on specific neurodevelopmental profiles in Kenya (KE) and South Africa (SA). This study is nested within NeuroDev, a larger case-control study which aims to characterise the genetic and phenotypic architecture of NDDs such as intellectual disability and autism in African populations.

(continue to page 81)

(continued from page 80)

This study included 490 children with NDDs, 31 affected siblings, and 255 controls. Data on socio-demographic factors were collected via questionnaires, and detailed medical and developmental assessments were conducted. Samples were collected for whole exome sequencing, which was performed on all participants. Single nucleotide variant (SNV) calling was done using the Genome Analysis Toolkit HaplotypeCaller, and copy number variant (CNV) calling followed GATK-gCNV. Logistic regression will be used for unadjusted and adjusted multivariable analysis of pre- and perinatal factors. We will describe known genomic disorders identified based on SNV/CNV findings, frequencies of rare variant rates, and associations between genetic findings and behavioural, cognitive, and medical phenotypes in KE and SA children. Among case families, 104 out of 483 had likely pathogenic or pathogenic SNV/CNVs. Preliminary unadjusted regression analysis found that childhood malaria exposure (OR:11.64 (95%CI 1.54 – 87.93)), fever during pregnancy (OR:8.72 (95%CI 2.60 – 28.57)), labour complications (OR:8.29 (%CI 3.21 – 21.36)) and having a sibling with an NDD (OR:4.88 (95%CI 3.74 – 6.14)) were associated with an increase in odds of NDDs in the KE sample. Male sex (OR:5.05 (95%CI 3.07 – 8.31) and being small for gestational age (OR:1.74 (95%CI 1.20 – 2.53)) were associated with NDDs in the SA sample. We hypothesize that a higher rare variant rate burden would be associated with more severe developmental outcomes and variable behavioural profiles. Results will be presented on rare variant burden and the characterization of genetic diagnoses, autism, and neurodevelopmental phenotypes in children of African ancestry from KE and SA.

Real-time dynamic polygenic prediction for streaming data

Abstract
Selected
Talk

Justin D. Tubbs[1,2,3], Yu Chen[3,4,5], Rui Duan[6], Hailiang Huang[3,4,5], Tian Ge[1,2,3]

1 Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA
2 Center for Precision Psychiatry, Department of Psychiatry, Massachusetts General Hospital, Boston, MA
3 Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA
4 Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA
5 Department of Medicine, Massachusetts General Hospital, Boston, MA
6 Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA

Polygenic risk scores (PRSs) hold promise in advancing precision medicine. However, existing PRS construction methods rely on static summary statistics derived from genome-wide association studies (GWASs), which are updated at lengthy intervals. As genetic data and health outcomes are continuously being generated at an ever-increasing pace, the current PRS training and deployment paradigm is suboptimal in maximizing the prediction accuracy of PRSs for incoming patients in healthcare settings.

We introduce real-time PRS-CS (rtPRS-CS), which uses a highly efficient stochastic gradient descent algorithm to update genome-wide single nucleotide polymorphism (SNP) weights within seconds as each new sample is collected, enabling online, dynamic refinement and calibration of PRSs. Through extensive simulation studies, we demonstrate that rtPRS-CS is robust to various genetic architectures and training sample sizes.

Leveraging 21 quantitative traits measured in the Mass General Brigham Biobank (MGBB) and UK Biobank (UKBB), we show that rtPRS-CS can efficiently integrate massive streaming data to enhance PRS prediction over time. For instance, when applying rtPRS-CS to body mass index (BMI) with initial SNP weights trained on a GWAS of 27,582 MGBB participants, the variance explained in BMI increases from a baseline of 5% by an average of 0.3% for each additional 10,000 UKBB samples incorporated. In an independent holdout sample, across all traits examined, rtPRS-CS significantly outperforms the PRSs constructed from baseline GWAS, with its predictive performance statistically indistinguishable from the theoretical upper limit.

Retrieval augmented generation greatly improves the ability of large language models to precisely assign human phenotype ontologies to clinical data

POSTER 405

Brandon T. Garcia[1], Priya Yelemali[1], Nikhita Gogate[1], E. Andres Rivera-Munoz[1], Angad Jolly[1], Moez Dawood[1], Haowei Du[1], Jennifer E. Posey[1]

1 Baylor College of Medicine, Department of Molecular and Human Genetics, Houston, TX

Accurate deep phenotyping of patient clinical data is essential for unraveling the complex relationship between genotype and phenotype in rare genetic diseases and for understanding the functionality of the protein-encoding genes of the human genome. Central to these goals and the mission of the Genomic Research to Elucidate the Genetics of Rare (GREGoR) Consortium is the use of Human Phenotype Ontology (HPO) terms, which provide a contextualized, standard vocabulary essential for both human and machine interpretation of phenotypic data. However, the need for refined phenotyping tools applicable to an international referral cohort lacking a uniform, standardized electronic medical records (EMR) source remains unaddressed.

To address this need, we developed RAG-HPO, a novel application that uses Retrieval Augmented Generation (RAG) and locally stored Large Language Models (LLMs) to accurately assign Human Phenotype Ontology (HPO) terms from medical free text. This user-friendly tool uses pre-processing techniques and clever prompt engineering combined with RAG, utilizing a shareable, synthetic database built from HPO project data and publicly available clinical case reports to quickly and precisely identify clinical findings and align them with available HPO terms.

We evaluated RAG-HPO paired with Llama3-70b on published clinical cases covering a wide variety of clinical conditions. Compared to a manually annotated standard and outputs from Doc2HPO, ClinPhen, and Llama-70b alone, RAG-HPO demonstrated substantial improvement in precision (0.85) and recall (0.73), with an F1 score of 0.78. In contrast, Doc2HPO achieved a precision of 0.67, recall of 0.40, and F1 score of 0.49, while ClinPhen showed a precision of 0.62, recall of 0.36, and F1 score of 0.43. Llama-70b performed poorly at baseline, with a precision of 0.18, recall of 0.13, and F1 score of 0.15.

RAG-HPO is an easy-to-use tool that ensures precise evaluation of medical free text for clinical and scientific users with minimal training in LLM usage. With its increased precision and recall, RAG-HPO expedites the exploration of genetic mechanisms underlying rare diseases. RAG-HPO enables fast and precise assignment of HPO terms, making it a significant asset for researchers aiming to expedite the identification of genotype-phenotype correlations in rare diseases or other applications in genomic medicine.

Retrieval augmented generation large-language model pipeline for clinical data abstraction from unstructured medical notes

POSTER 406

Thinh Tran[1], Noah Friedman[1], Miller Richters[1], J. Bryce Ortiz[1], Richard Green[1], Robert Burns[1], Helio Costa[1]

1 Natera, Inc., Austin, TX

Background: In clinical genetic testing, patient health records contain a wide variety of clinical information of high value. However, the data is often contained within unstructured clinical notes, making it challenging to encode it into a clinico-genomic database. Here, we leveraged advances in natural language processing, specifically Large Language Models (LLMs), to develop a pipeline that could comprehend and format heterogeneous and unstructured textual medical records.

Methods: We designed a pipeline that ingested unstructured medical records and extracted specified clinical outputs using LLM in a HIPAA-compliant manner. Optical character recognition (OCR) was utilized to convert scanned documents into machine-readable text. Following OCR conversion, a retrieval augmented generation pipeline was implemented that used Facebook AI Similarity Search to retrieve sections of the patient records relevant to desired clinical attributes. The retrieved snippets, along with a series of prompts, were used together to extract desired output using a LLM. A series of prompts were used to prepare the structured clinical data comprising a primary "summary" prompt to summarize the relevant clinical information, and a second "formatting" prompt to transform the one-paragraph summary into a structured JSON format using Anthropic's Claude (v2.1).

Results: We evaluated LLM performance on abstracting medication names from over 100 unstructured medical records. The LLM-abstracted attributes were compared with manually-abstracted attributes from the same set of medical records by trained clinical data abstractors. Our results indicated that the LLM was highly accurate in abstracting medication names, achieving an accuracy of 0.63, precision of 0.66 and recall of 0.86 when compared against manually-abstracted data. The LLM pipeline is highly scalable and can be parallelized to abstract attributes from thousands of medical records in one run.

Conclusions: These results demonstrate the clear operational value of LLMs for streamlining the process of clinical data abstraction into a large clinico-genomic database. Widespread use of these tools has the potential to allow for a rapid, cost and time-efficient means to comprehend clinical information for scientific advances and ultimately improved patient care.

Revolutionizing epigenomic analysis in human disease: high-resolution spatial CUT&Tag and spatial ATAC-seq mapping at the single-nucleus level

POSTER 407

Katelyn Noronha[1], Silas Decker[1], Gumaro Rojas[1], Jennifer Garbarino[1], Molly Wetzel[1], Machele Riccio[1], Jose Perez[1], Abigail Tyree[1], Arianna McDaniels[1], Jeffrey Sabina[1], Colin Ng[1]

1 AtlasXomics Inc., New Haven, CT

Advances in spatial transcriptomics and proteomics have enabled increasingly complex investigations into gene expression across various cellular subtypes and tissues. However, these methods do not explore the epigenome, which regulates gene expression in a cell-specific manner and plays a critical role in human diseases including cancer, inflammation, and neurodegeneration. A nuanced understanding of epigenetic dysregulation of chromatin structure and regulatory proteins can drive therapeutic development. To address the need for tools beyond bulk molecular assays for this research, we have enhanced the application of deterministic barcoding in tissue for spatial omics sequencing (DBiT-seq), a technique first published in 2022 by Dr. Rong Fan's lab at Yale. Our platform extends the assay area and resolution to investigate a 5.5 mm x 5.5 mm region at the single-nucleus level. This platform enables spatial cleavage under targets and tagmentation (spatial CUT&Tag), allowing for the detection of both histone modifications and chromatin-associated proteins in tissues and revealing complex spatial gene regulation patterns. The platform can also map open chromatin in tissues using spatial ATAC-seq. We demonstrate detection of cellular subtypes and regulatory elements at specific genes important to tumor biology by spatial chromatin accessibility and activating histone mark profiling in human gastric adenocarcinoma, adjacent normal tissue, and adjacent lymph node. We have also developed a fully integrated bioinformatics suite designed to allow non-bioinformatics users to easily explore and identify spatial epigenetic elements. Furthermore, we have created a framework that enables easy integration with existing spatial and single-cell datasets for a unique multi-omics solution necessary for the complexity of the tissue environment in disease. This innovative platform not only paves the way for unprecedented insights into the epigenetic landscape but also holds the potential to transform our approach to personalized medicine and therapeutic intervention.

SimPheny: Matching similar phenotypes across rare disease cohorts for genetic diagnosis

POSTER 408

Isabelle Cooperstein[1], Alistair Ward[1,2], Barry Moore[1], Gabor T. Marth[1]

1 University of Utah, Department of Human Genetics, UT
2 Frameshift Labs, Inc., Cambridge, MA

Patients with rare monogenic diseases often face prolonged diagnostic odysseys due to the challenge of pinpointing a molecular diagnosis among millions of genomic variants. Computational approaches are essential for identifying functionally relevant variants for review. However, current variant prioritization tools are limited by incomplete knowledge of phenotype-to-gene associations and the complex, diverse nature of clinical phenotypes, which often diverge from known disease descriptions. We introduce SimPheny, a novel approach that automates the identification of patients with similar phenotypes to prioritize genes for diagnostic review, and demonstrate its effectiveness using the Undiagnosed Diseases Network (UDN) cohort. SimPheny utilizes Human Phenotype Ontology (HPO) terms to calculate pairwise phenotypic similarity scores between individuals. We benchmarked SimPheny using a cohort of UDN patients with known diagnostic variants and whole genome sequencing data. Exomiser was used to generate permissive gene lists of potential disease-causing variants, which SimPheny prioritized by identifying overlaps with diagnostic genes of phenotypically similar patients, termed "gene hits." We developed statistical methods to evaluate the significance of each hit and the false discovery rate, establishing thresholds to identify likely diagnostic gene hits. SimPheny was then applied to prioritize genes in 1,319 as-yet unsolved UDN cases.

Our method effectively measures high phenotypic similarity between individuals who have molecular diagnoses in the same genes within the clinically diverse UDN cohort. In the benchmarking study, Exomiser ranked the true diagnostic gene in 92% of cases, but with an average of 185 genes per individual. SimPheny significantly reduced the number of potential diagnostic genes per patient, achieving a manageable list of between 1 and 8 genes at varying thresholds with high precision rates. Applying SimPheny to the undiagnosed UDN cohort yielded numerous significant hits that are currently under investigation, showing promise for new diagnoses.

SimPheny leverages phenotypic and genomic data to automate patient matching, demonstrating its effectiveness in gene prioritization within the UDN. This method offers a promising path to improve diagnostic rates and uncover new gene associations in rare monogenic diseases. SimPheny is now publicly available as an iobio web application, primed for seamless integration into diagnostic workflows.

Spatial molecular analysis of human pancreatic tissues based on P-PASS parameters

POSTER 409

Megan Pater[1], Grant R. Kolar[1], Joshua D. Stafford[1], Willis K. Samson[1], Gina L.C. Yosten[1]

1 Saint Louis University School of Medicine, Department of Pharmacology and Physiology, Saint Louis,

Type I diabetes (T1D) is an autoimmune disease characterized by the destruction of insulin-secreting pancreatic beta cells. While insulin replacement therapy is the most common treatment, pancreas transplantation is the only current curative option for T1D. The need for pancreas transplantation has increased over the years, yet the number of tissues available for transplantation has decreased. Eligibility of a pancreas for donation is determined by P-PASS criteria that includes age and BMI. However, studies have shown tissues that marginally pass P-PASS criteria demonstrate high success rates post-transplantation suggesting that a broadening of the P-PASS criteria may increase the number of successful transplants. In this study, we aimed to compare the spatial transcriptomic profiles of human pancreatic tissue across P-PASS score, age and BMI. We sought to identify differentially expressed genes that could potentially serve as new biomarkers for pancreas transplant or homology in transcriptional profiles between P-PASS criterion groups. Thus, potentially leading to changes in the P-PASS parameters. Along with P-PASS criteria, pancreases may be rejected by a surgeon based on appearance, however, as the pancreas is largely comprised of exocrine tissue, this rejection criteria weighs heavily on cell types not related to insulin secretion. Thus, we also determined the influence of the transcriptomic profile of acinar cells surrounding the Islet of Langerhans on the transcriptional profiles of the endocrine tissue.

For this study, 183 patient pancreata were collected via MidAmerica Transplant Center (Saint Louis, MO) and analyzed using the 'Human Universal Cell Characterization panel' [1000-plex] by the Nanostring CosMx Spatial Molecular Imager. We present initial results which demonstrate stark differences in the differential transcriptomic profiles of pancreata when stratified based on the BMI and age P-PASS parameters and therefore, cannot be combined. This study is the foundation for our attempts to find potential biomarkers or new criteria to determine suitability of pancreases for transplantation.

Spotiphy enables single-cell spatial whole transcriptomics via generative modeling

POSTER 410

Jiyuan Yang[1], Ziqian Zheng[2], Yun Jiao[3], Kaiwen Yu[4], Sheetal Bhatara[1], Xu Yang[1], Sivaraman Natarajan[1], Jiahui Zhang[2], John Easton[1], Koon-Kiu Yan[1], Junmin Peng[3], Kaibo Liu[2], Jiyang Yu[1,5]

1 Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN
2 Department of Industrial & Systems Engineering, University of Wisconsin-Madison, Madison, WI
3 Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, TN
4 Center of Proteomics and Metabolomics, St. Jude Children's Research Hospital, Memphis, TN
5 Lead contact.

Spatial transcriptomics (ST) has revolutionized our understanding of tissue regionalization by making it possible to visualize gene expression across an intact tissue section, but the technical limitations of existing ST approaches leave an unmet need for single-cell-resolved transcriptomic profiles with genome-wide coverage. Here we present Spotiphy (Spot imager with pseudo single-cell resolution histology), a novel algorithm that resolves the "gene coverage-resolution" tradeoff faced by current sequencing- and image-based ST platforms and transforms sequencing-based ST data into single-cell-resolved whole-transcriptome images. It achieves this by (i) leveraging generative modeling with single-cell RNA sequencing (scRNA-seq) data and high-resolution histological images to convert spot-level ST data into single-cell whole transcriptomic profiles, (ii) utilizing Gaussian processes to impute the cellular composition and expression profiles of non-capture areas, and then (iii) merging the results of the first two steps into a whole-slide image. In evaluations that use matched scRNA-seq, Visium, Xenium, CosMx, and immunohistochemistry (IHC) datasets for Alzheimer's Disease (AD) and normal mouse brains, Spotiphy delivers the most precise cell-type proportions and accurately depicts the distribution of rare cell populations including immune cells. Applying computational workflows to the single-cell profiles generated by Spotiphy reveals novel astrocyte regional specification in mouse brains, as validated by image-based CosMx results. Furthermore, Spotiphy distinguishes DAM (Disease-Associated Microglia) sub-populations located in different AD mouse brain regions. Applied to ST data of human breast samples, Spotiphy identifies multiple spatial domains as well as changes in the patterns of tumor-tumor microenvironment (TME) interaction across these domains.

(continue to page 89)

(continued from page 88)

For the first time, Spotiphy tackles the information loss at the non-capture area using Gaussian processes and provides superior cell-type proportions and single-cell expression profiles with high accuracy validated by matched ground truth. It delivers a single-cell-resolved whole-transcriptome image of the entire section that significantly expands the information intensity of sequencing-based ST data, providing a more comprehensive and detailed understanding of spatially resolved cellular states and regulatory mechanisms. By making it possible to visualize the cellular localization and expression profiles of an intact tissue section, Spotiphy will be an important tool for gaining insight into the cellular organization, tissue heterogeneity, and function of complex biological systems.

STAT5A activation caused lipid dysregulation in FLNC deficient cardiomyocyte.

POSTER 411

Shanshan Gao[1], Lingaonan He[1], Suet Nee Chen[1], Matthew R.G. Taylor[1], Luisa Mestroni[1]

1University of Colorado Anschutz Medical Campus, Division of Cardiology, Aurora, CO

Background: Truncating variants in Filamin-C (FLNC) gene are associated with dilated cardiomyopathy (DCM), but the molecular mechanisms underlying pathogenesis of FLNC-related cardiomyopathy remains largely unknown. Lipid droplets (LDs) are dynamic organelles which store fat in the form of neutral lipids, mainly triglyceride (TAG). LDs play an essential role in the regulation of lipid metabolism and serve as hubs for energy homeostasis. Nevertheless, excessive retention of LDs is commonly observed in the hearts of DCM patients, but whether FLNC mutation contributes to the pathogenic alteration of cardiac LDs in DCM has not been elucidated. Thus, the aim of our study is to unravel the key signaling pathways involved in cardiac lipid dysregulation in DCM caused by FLNC truncating (FLNCtv) mutation.
Methods and results: We have generated FLNC-knockout iPSC edited with the CRISPR/Cas system and cardiac myocytes specific knockout mice. We found increased formation of LDs in FLNC-knockout induced pluripotent stem cells derived cardiomyocytes (iPSC-CMs) as compared to control iPSC-CMs. Moreover, we observed depressed cardiac function with ventricular dilatation in cardiomyocyte specific Flnc knockout, Myh6-Cre:FLNCW/F mice. Notably, transmission electron microscopy imaging revealed significantly more LD accumulation in the myocardium of Myh6-Cre:FLNCW/F mice than in the hearts of non-transgenic animals supporting our observation in FLNC mutant iPSC-CMs derived from patients with FLNCtv mutations. Moreover, we found that Tyr-694 phosphorylation of STAT5A, a transcription factor that mediates adipogenesis, was significantly increased in FLNC-knockout iPSC-CMs. In addition, phosphorylated STAT5A expression levels were upregulated in the hearts of the Myh6-Cre:FLNCW/F mice compared to the wildtype (WT) littermate controls. Furthermore, STAT5A phosphorylation levels were significantly increased in the explanted hearts of FLNCtv carriers with DCM compared to non-failing hearts. We then treated FLNC knockout iPSC-CMs and adult cardiomyocytes isolated from Myh6-Cre:FLNCW/F mice with AC-4-130, a potent STAT5 inhibitor and showed that lipid droplets were reduced in both FLNC-knockout iPSC-CMs and cardiomyocytes derived from Myh6-Cre:FLNCW/F mice.
Conclusion: In conclusion, these results strongly suggest that STAT5A activation contributes to lipid metabolism dysregulation in DCM caused by FLNC truncating variants.

STRchive: a dynamic resource detailing population-level and locus-specific insights at tandem repeat disease loci

POSTER 412

Laurell Hiatt[1], Ben Weisburd[2], Egor Dolzhenko3, Grace VanNoy[2], Nehir Kurtas[4], Heidi Rehm[4], Aaron Quinlan[5], Harriet Dashnow[5]

1 University of Utah, Salt Lake City, UT
2 Broad Institute of MIT and Harvard, Cambridge, MA
3 Pacific Biosciences, Menlo Park, CA
4 Massachusetts General Hospital, Boston, MA
5 University of Utah, Salt Lake City, UT

Approximately 3% of the human genome consists of repetitive elements called tandem repeats (TRs), which include short tandem repeats (STRs) of 1–6bp motifs and variable number tandem repeats (VNTRs) of 7+bp motifs. TR variants contribute to several dozen mono- and polygenic diseases but remain understudied and "enigmatic," particularly relative to single nucleotide variants. It remains comparatively challenging to interpret the clinical significance of TR variants. Although existing resources provide portions of necessary data for interpretation at disease-associated loci, it is currently difficult or impossible to efficiently invoke the additional details critical to proper interpretation, such as motif pathogenicity, disease penetrance, and age of onset distributions. It is also often unclear how to apply population information to analyses.

We present STRchive (S-T-archive, http://strchive.org/), a dynamic resource consolidating information on TR disease loci in humans from research literature, up-to-date clinical resources, and large-scale genomic databases, with the goal of streamlining TR variant interpretation at disease-associated loci. We apply STRchive —including pathogenic thresholds, motif classification, and clinical phenotypes—to a gnomAD cohort of ~18.5k individuals genotyped at 60 disease-associated loci.

Through detailed literature curation, we demonstrate that the majority of TR diseases affect children despite being thought of as adult diseases. Additionally, we show that pathogenic genotypes can be found within gnomAD which do not necessarily overlap with known disease prevalence, and leverage STRchive to interpret locus-specific findings therein. We apply a diagnostic blueprint empowered by STRchive to relevant clinical vignettes, highlighting possible pitfalls in TR variant interpretation. As a living resource, STRchive is maintained by experts, takes community contributions, and will evolve as understanding of TR diseases progresses.

Systems genetics of metabolic health in the BXD mouse genetic reference population

POSTER 501

Xiaoxu Li[1], Jean-David Morel[1], Jonathan Sulc[1], Alessia De Masi[1], Amélia Lalou[1], Giorgia Benegiamo[1], Johanne Poisson[1], Yasmine

1 Laboratory of Integrative Systems Physiology, Institute of Bioengineering, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
2 Laboratory Genetic Metabolic Diseases, Amsterdam Gastroenterology, Endocrinology and Metabolism Institute, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, The Netherlands

Susceptibility to the metabolic syndrome (MetS) is dependent on genetics, environment and gene-by-environment interactions, rendering the study of underlying mechanisms challenging. The majority of experiments in model organisms do not incorporate genetic variation and lack specific evaluation criteria for the MetS. Here, we derived a continuous metric, the metabolic health score (MHS), based on standard clinical parameters and defined its molecular signatures in the liver and circulation. In human UK Biobank, the MHS associated with the MetS status and was predictive of future disease incidence even in individuals without the MetS. Using quantitative trait locus analyses in mice, we found two MHS-associated genetic loci and replicated them in unrelated mouse populations. Through a prioritization scheme in mice and human genetic data, we identified TNKS and MCPH1 as candidates mediating differences in the MHS. Our findings provide insights into the molecular mechanisms sustaining metabolic health across species and uncover likely regulators.

Transcriptomic profiling and isoform switching as biomarkers for seizure type differentiation

POSTER 502

Rashi Verma[1], Katie Bullinger[2], Andrea Pearson[1], Monica Dhakar[2], Emine Guven[1], Elham Amini[1], Robert Meller[1,3*]

1 Neuroscience Institute, Morehouse School of Medicine, Atlanta, GA
2 Department of Neurology, Emory University, Atlanta, GA
3 Institute of Translational Genomic Medicine, Morehouse School of Medicine, Atlanta, GA

Objective: Accurate diagnosis of seizure types is essential for personalized treatment in epilepsy management. Building on prior findings that RNA signatures can distinguish non-epileptic from epileptic seizures, this study investigates alternative splicing of RNA transcripts to determine whether transcript usage can serve as precise biomarkers for differentiating seizure types.

Methods: Blood samples from 27 patients undergoing video electroencephalogram (vEEG) monitoring at Emory University Hospital were collected at baseline, 4-6h post-seizure, and discharge. Epileptologists determined seizure etiology through vEEG data review. RNA was extracted from blood samples for RNA sequencing. R packages were employed to analyze RNA expression and transcript usage. Prediction models were generated using the caret package to distinguish between patients adjudicated to have had a focal or generalized seizure.

Results: Distinct transcriptomic profiles were observed between focal and generalized seizure across time points, with the most pronounced differences at baseline and discharge. Patients with focal seizures showed downregulation of transcripts at 4-6h post-seizure and upregulation at discharge, while generalized seizure patients exhibited opposite pattern. Notably, significant isoform switching (n = 1249 genes) without differences in gene expression was viewed ($|dIF| > 0.1$). Intron retention and alternative transcript termination sites were identified as key biological consequences of isoform switching. The sensitivity to nonsense-mediated mRNA decay (NMD) was assessed for the identified isoforms, revealing 742 NMD-sensitive genes. Among these, RNF216, with its novel NMD-sensitive transcript (MSTRG.52862.7), emerged as a top gene exhibiting differential transcript usage across seizure types. Furthermore, differentially expressed transcripts were utilized as features for machine learning classification. Remarkably, models generated using random forest and radial support vector machine algorithms achieved ~83% accuracy in predicting generalized seizures, while the multivariate adaptive regression splines algorithm achieved 100% accuracy in predicting focal seizure events.

(continue to page 94)

(continued from page 93)

Significance: This study provides the first comprehensive analysis of transcriptome expression, specifically focusing on isoform switching, to differentiate seizure types. The results highlight potential biomarkers, promising advances in epilepsy diagnosis and therapy.

Uncovering brain vascular transcriptional changes across age and Alzheimer's disease

POSTER 503

Alex J. Pennacchio[1], Madigan Reid[1*], Tammie Tam[2*], Bella Ding[1], Brittany Davidson[2], Jessica Tsui[2], Amanda Apolonio[1], Hao Liu[1], Shih-Hsiu J. Wang[3], Alexis Combes[2,#], Andrew C. Yang[1,2,#]

1 Gladstone Institutes, Gladstone Institute of Neurological Disease, San Francisco, CA
2 University of San Francisco, Department of Pathology, San Francisco, CA
3 Department of Pathology, Duke University Medical Center, Durham, NC
*Contributed equally
#These authors jointly supervised this work

The blood-brain barrier (BBB) is a selectively permeable protective structure that regulates the transport of necessary molecules while protecting the brain from harmful substances. Formed by a specialized community of vascular and immune cells, the BBB is essential for maintaining healthy brain function. However, in the process of aging, especially in individuals developing Alzheimer's disease (AD), key BBB functions become disrupted, evidenced by increased leakage of blood components via histopathology and clinical imaging. Brain vascular dysfunction is now recognized as one of the earliest pathological changes associated with late-onset AD. For example, disruptions in cerebral blood flow and damage to the BBB can precede clinical symptoms by decades. However, molecular changes to the BBB in aging and disease are not well characterized, in part due to technical shortcomings of traditional scRNA-seq workflows in detecting functionally important vascular cell types.

The recent development of vessel isolation and nuclei extraction for sequencing (VINE-seq), enables large-scale capture of vascular cell types, which have been under-represented in previous AD scRNA-seq data. In this study, we generate snRNA-seq data in paired cortical gray and adjacent white matter samples from a cohort of 105 individuals. Our cohort is composed of post-mortem healthy brains of a wide range of ages, aged AD brains, and aged resilient brains, which have AD-like amyloid-beta plaque pathology but do not display cognitive decline. We captured over 1,000,000 single cells, including more than 150,000 vascular cells, which we leveraged for downstream analyses. We are currently conducting differential expression analysis across age and disease groups to identify genes that contribute to AD development versus healthy aging, with a focus on vascular cell types. Additionally, we are employing 10x Xenium spatial transcriptomics to understand how vascular cell types change expression in situ locally around AD pathologies and corresponding changes in immune cell distribution and activation state. This study offers value in the size and diversity of the cohort and emphasis on vascular and immune contributions to aging and AD.

Utilizing Targeted Enhanced-Whole-Genome Sequencing in Precision Oncology for the Treatment of Solid Tumors: A Clinical Perspective

POSTER 504

Erin Connolly-Strong[1] Amy Ackershoek[1], Alexis Barcomb[1], Brian Baek-Lok Oh[1], Chunyang Bao[1], Jung Ah Kim[1], Eun Kyoung Lee[1], Jeong-Hoon Lee[1], Sangmoon Lee[1]

1 Inocras Inc., San Diego, CA

Background: Understanding the genomic landscape is essential for effective cancer management and treatment. Targeted panel sequencing and whole exome sequencing have been key in identifying known mutations but only cover limited genetic regions, potentially missing significant insights. Recent advancements and cost reductions in whole-genome sequencing (WGS) provide a comprehensive alternative, examining the entire genomic landscape and capturing both common and rare mutations, including those in non-coding regions that may influence cancer management. This shift enhances precision cancer treatment by offering a more complete genetic profile, paving the way for personalized medicine.
Methods: We conducted a retrospective analysis of 52 cases using target-enhanced whole-genome sequencing (TE-WGS) by Cancer Vision (Inocras, San Diego, CA). The TE-WGS assay, validated for comprehensive oncologic genomic profiling, demonstrated 99.8% sensitivity for single nucleotide variants(SNVs), 99.2% for insertions/deletions(indels), and positive predictive values of 99.3% for SNVs, and 98.7% for indels. This study assessed the clinical utility of TE-WGS in informing targeted therapy selection, clinical trial opportunities, and diagnostic clarity.
Results: This pan-cancer study included various disease states, such as breast, colorectal, lung, and sarcoma. TE-WGS demonstrated clinical utility in 65% of evaluated cases. Notably, 42% of these cases revealed actionable genomic alterations that aligned with on-label and/or off-label targeted therapies, and 44% of patients exhibited genomic profiles matching ongoing clinical trials.
Conclusion: TE-WGS significantly enriches the genomic landscape accessible to clinicians, which enhances personalized treatment strategies in precision oncology through comprehensive and actionable genomic profiling.

Validating third-generation Isoform sequencing data using first-generation and second-generation sequencing technologies

POSTER 505

Chioniso Patience Masamha[1], Mahesh Gupta[1], Caiden Lukan[1], Lael Pasipamire[1,2], Jie Li[3]

1 Butler University, Pharmaceutical Sciences, College of Pharmacy and Health Sciences, Indianapolis, IN
2 Purdue University, Biomedical Sciences, Lafayette, IN
3 University of California Davis, UC Davis Genome Center, Davis, CA

There are several mechanisms that give rise to transcript diversity in human cells. These include alternative promoter usage, alternative splicing, alternative polyadenylation, gene fusions, and transcriptional readthrough events that give rise to fusion transcripts. In the past it has been difficult to globally identify all these different transcripts simultaneously. Here we used third generation sequencing technology, Iso Sequencing, on the hematopoietic malignancy mantle cell lymphoma. After bioinformatic analysis, we identified a wide repertoire of different transcripts including known and novel fusion transcripts. We will present our results using one representative fusion transcript, two alternatively spliced transcripts, and transcripts arising from alternative polyadenylation within the same terminal exon of the same gene. We will discuss how we developed primers and used PCR, cloning and Sanger sequencing to validate these fusion transcripts. In addition, we used both Iso Sequencing and RNA Sequencing to identify alternatively spliced transcripts. Here we will demonstrate how we used this data to identify the two alternatively spliced isoforms and used qRT-PCR and Western blots to quantify their levels and their expression at the protein level. Finally, we will discuss how we validated the presence of alternatively polyadenylated transcripts using primers we developed for both qRT-PCR and 3'RACE. We will also discuss how the alternatively polyadenylated transcripts were cloned and sequenced using Sanger Sequencing to validate their presence in the mantle cell lymphoma samples. Full understanding of the transcriptome of human cancers is important because it will provide biomarkers for disease diagnosis and patient stratification, as well as provide potential therapeutic targets using RNA based technologies.

VariLink™: a novel assay for comprehensive variant detection and genome characterization and its comparison to WGS

POSTER 506

Ericca Stamper[1], Jon Torchia[1], Daniel Hwang[1], Chu Yuan Wu[1], Ek Danwanichakul[1], Cory Padilla[1], Lisa Munding[1], J. Zachary Sanborn[1]

1 Dovetail Genomics, Part of Cantata Bio, LLC, Scotts Valley, CA

Many diseases are caused by genetic variation. Currently, no single technology can capture the wide range of genetic variation with high sensitivity. We developed VariLink™ – a multi-purpose proximity ligation assay that yields high-quality, unbiased Hi-C libraries that capture not only chromatin conformation linked to epigenetic state, but are capable of highly sensitive discovery of structural variants (SVs). Importantly, VariLink data displays uniform coverage and thus enables similar performance to whole genome shotgun sequencing (WGS) for the detection of single nucleotide variants (SNVs), insertions / deletions (InDels), and copy number variants (CNVs). VariLink libraries can be generated in a single day and are fully compatible with Illumina sequencers.

To assess the real-world performance of VariLink relative to standard NGS methods, we selected a clinical ovarian serous adenocarcinoma for genomic analysis using VariLink and WGS to detect SVs, SNVs, InDels, and CNVs, as well as RNA-seq to find gene fusions induced by SVs. VariLink and WGS libraries were sequenced to a ~30x coverage. RNA-seq libraries were sequenced to yield over 100M reads.

VariLink data discovered three translocations (tlocs) between chromosomes 1, 3, 4, and 12 with strong supporting evidence (569 - 3,232 read pairs). These tlocs were supported by changes in copy number and were located at CNV boundaries. Furthermore, all 3 tlocs were validated by PCR and Sanger sequencing. By comparison, only one fusion was detected by RNA-seq, while WGS detected 2 out of 3 tlocs with low precision (ranked 538 and 946 of 1,018 putative tlocs sorted by decreasing read pair support).

(continue to page 99)

(continued from page 98)

We find VariLink outperforms WGS and RNA-seq in SV discovery with both higher sensitivity and significantly higher precision. This result, coupled with VariLink's similar performance to WGS across other variant types, indicate VariLink is a powerful tool for genetic variation discovery.

Preimplantation genetic testing (PGT) is widely used in in vitro fertilization (IVF) to identify genetic defects in embryos before transfer. Currently, PGT requires a trophectoderm (TE) biopsy, which is invasive and subject to sampling bias. Non-invasive preimplantation genetic testing (ni-PGT) using cell-free DNA from spent embryo culture medium (SCM) and blastocoele fluid (BF) offers a promising alternative. However, current niPGT methods face accuracy challenges, especially for monogenic diseases. Challenges for niPGT-M include low DNA content, maternal contamination, and high allele drop-out (ADO) rates.

We developed a novel method integrating a modified linear single-cell whole genome amplification (sc-WGA) technique with a Bayesian linkage analysis model to improve the accuracy of non-invasive PGT for monogenic diseases (niPGT-M). Eleven patients undergoing ICSI and PGT-M with TE biopsy at Peking University Third Hospital were recruited for this study. Spent culture medium from Day 4 to Day 5/6 of each embryo was collected. Using the modified scWGA method, all 90 spent culture media samples were successfully amplified. Through the Bayesian linkage analysis method, samples with extremely low-quality data were filtered out. This new analysis method incrementally incorporates SNPs, allowing us to calculate the log-likelihood ratio of inheriting disease-carrying chromosomes versus disease-free chromosomes for each SNP subset. This enables classification into four categories: Highly Confident, Moderately Confident, Possible, and Undetermined. 74.8% of the samples were successfully classified into the first three categories, yielding results for niPGT-M. Compared to TE-biopsy PGT-M results, which are considered as the gold standard, the accuracy of niPGT-M was 100%, significantly reducing the risk of misdiagnosis compared to conventional linkage analysis methods.

Among the samples with euploid niPGT results, we successfully reconstructed the whole genome of embryos and calculated the Polygenic Risk Score (PRS) for type II diabetes in non-invasive preimplantation genetic testing for polygenic diseases (niPGT-P). This approach not only efficiently amplifies cell-free DNA but also greatly enhances the accuracy of analysis, advancing the potential clinical application of niPGT.

Whole genome sequence analysis for comprehensive discovery and functional prioritization of Mendelian risk variants

POSTER 507

Arthur S. Lee[1,2*], Alba Sanchis-Juan[1,2*], Eren Shin[1,2], Nehir E. Kurtas[1,2], Jack M. Fu[1,2], Xuefang Zhao[1,2], Justin Lim[1,2], Yulia Mostovoy[2], Gabrielle Lemire[2], Nicole Calamari[1,2], Michael Kosicki[3], GREGoR Consortium, Broad Center for Mendelian Genomics, GMKF OFC Working Group, Broad Genomics Platform, Angie Jelin[4], Terri H. Beaty[5], Frances A. High[6], Patricia K. Donahoe[6], Stephanie M. Ware[7], Ian D. Krantz[8], Joseph Gleeson[9], Azeez Butali[10], Elizabeth C. Engle[11], Ali Gharavi[12], Anne O'Donnell-Luria[1,2,13], Heidi L. Rehm[1,2], Wendy Chung[13], Elizabeth J. Leslie[14], Mary L. Marazita[15], Axel Visel[3], Len A. Pennacchio[3], Harrison Brand[1,2], Michael E. Talkowski[1,2]

1 Department of Neurology, Harvard Medical School, Boston, MA; Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA
2 Broad Institute Center for Mendelian Genomics, Broad Institute of MIT and Harvard, Cambridge, MA; Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA
3 Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA
4 Department of Gynecology and Obstetrics, The Johns Hopkins Medical Institutions, Baltimore, MD
5 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
6 Pediatric Surgical Research Laboratories, Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA
7 Department of Pediatrics and Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN
8 Division of Human Genetics, The Children's Hospital of Philadelphia, Perelman School of Medicine at University of Pennsylvania, Philadelphia, PA
9 Department of Neurosciences, University of California San Diego, La Jolla, CA
10 Department Oral Pathology, Radiology and Medicine, College of Dentistry, University of Iowa, Iowa City, IA
11 Department of Neurology and Ophthalmology, Boston Children's Hospital and Harvard Medical School, Boston, MA; Howard Hughes Medical Institute, Chevy Chase, MD
12 Center for Precision Genetics and Genomics, Department of Medicine, Columbia University, New York, NY
13 Department of Pediatrics, Boston Children's Hospital, Harvard Medical School, Boston, MA
14 Department of Human Genetics, Emory University School of Medicine, Atlanta, GA
15 Center for Craniofacial and Dental Genetics, Department of Oral and Craniofacial Sciences, University of Pittsburgh, Pittsburgh, PA

(continue to page 101)

(continued from page 100)

Rapid advances in sequencing technologies have empowered investigators studying congenital anomalies to generate a wealth of patient genomic data at a massive scale. Despite these advances, causal variant discovery, interpretation, and diagnosis remain far from a routine task, as >50% of rare disease cases receive no genetic diagnosis. Discovery remains limited by several key factors, including: (1) modest cohort sizes that restrict power to detect recurrent mutations; (2) failure to exhaustively capture and genotype major classes of disease-causing variation across the genome; and (3) our inability to interpret and subsequently validate candidate mutations at scale, particularly for noncoding variants. Therefore, to more comprehensively explore the distinct and shared genetic architecture of rare congenital anomalies, we aggregated patient sequencing data across eleven international consortia, totalling 20,993 trios (15,793 exomes and 5,200 genomes) with structural birth defects (SBDs). We performed joint genotyping, QC, and harmonization of SNVs/indels and structural variants (SVs) across all cohorts using GATK and GATK-SV, respectively. We performed association using the transmission and de novo association (TADA) framework, a Bayesian statistical model that combines evidence across variant classes, weighted by metrics of evolutionary constraint. We identified 217 genes with a false discovery rate < 0.01, 22 of which have never been associated with syndromic SBDs. Extending our analysis to noncoding variants, we developed an unbiased Nocoding Combinatorial Association Selection (NCAS) model to estimate the effects of purifying selection of SVs overlapping 198,720 unique combinations of noncoding annotations. Leveraging SVs from the gnomAD database, we identified 616 Bonferroni-corrected significant ($p < 4x10E-06$) combinations implicating potentially disease-relevant classes of noncoding variation (e.g., deletions overlapping conserved DNase I hypersensitive sites, TAD boundaries, and lncRNAs). Finally, we benchmarked our observed versus expected SV estimates using the foundational DevCisReg catalogue, a database of noncoding elements that are active during embryonic development. Interestingly, we observed a significant global depletion of rare gnomAD SVs across DevCisReg elements ($p < 10E-100$), particularly for constrained elements in the brain. Altogether these analyses deliver unbiased, testable loci for downstream functional validation and underscore the untapped potential of leveraging aggregated genomic data across diverse disease cohorts.

Whole genome sequencing for the comprehensive clinical evaluation of B-lymphoblastic leukemia

POSTER 508

Eric Duncavage[1], Andrew Hughes[1], Feiyu Du[2], Nidhi Davarapalli[3], Shelly O'Laughlin[2], Molly Schroder[1], Geoff Uy[3], David Spencer[3]

1 Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO
2 Department of Genetics, Washington University School of Medicine, St. Louis, MO
3 Department of Medicine, Division of Oncology, Washington University School of Medicine, St. Louis, MO

Intro:
B-acute lymphoblastic leukemia (B-ALL) accounts for 6,500 new cancer cases a year. The classification of B-ALL is now based exclusively on genomic findings and the 2022 International Consensus Classification (ICC) recognizes 22 unique B-ALL genomic subtypes based on the detection gene rearrangements, single gene mutations, gene expression signatures, and DNA ploidy. We sought to determine whether a single clinical whole genome sequencing (WGS) assay, ChromoSeq, could be used to classify B-ALL using 2022 ICC guidelines.

Methods:
WGS was performed on initial diagnostic bone marrow DNA from 74 B-ALLs (9 pediatric cases and 65 adults) using Illumina DNA library preparation and NovaSeq X+ sequencing with 2x150bp reads. Data was analyzed using the Illumina Dragen 4.3 software (which includes improved identification of DUX4 rearrangements) followed by a custom analysis workflow designed to filter known variants. WGS data was compared to existing clinical data including cytogenetics, FISH, targeted NGS data if available.

Results:
WGS was performed to a mean depth of 52.9x. Comprehensive analysis using ChromoSeq assigned 66 of 74 cases (89%) to a specific 2022 ICC genomic category compared to 41 cases (55%) using existing clinical data. BCR::ABL1 rearrangements were correctly identified in 20% of cases (15/74). The second most commonly encountered category was BCR::ABL1–like, JAK-STAT activated, which accounted for 18% of cases (13/74), including 11 with rearrangements involving CRLF2, 1 EPCOR rearrangement, and 1 PAX5:JAK2 rearrangement, all of which were confirmed by RNAseq. Other rearrangements (ETV6, KMT2A, MEF2D, MYC) were detected in 27% of cases (20/74). Ploidy changes were detected in 22% of cases (16/74), and PAX5 p.P80R and a BCR::ABL1–like, ABL-1 class rearrangement were each identified in 1 case.

(continue to page 103)

(continued from page 102)

Conclusions:
We demonstrate that a single, clinically optimized, tumor-only WGS assay, ChromoSeq, is capable of providing a 2022 ICC genomic classification in 89% of B-ALL cases. WGS-based B-ALL classification has the potential to streamline diagnostic workflows that currently rely on multi-modality testing and could decrease turnaround times and lower overall testing costs. Further, the use of a standardized approach such as WGS would simplify the comparison of genomic findings across centers, facilitating enrollment in clinical trials.

Whole genome sequencing in an adolescent idiopathic scoliosis cohort indicates polygenic disease involving multiple biological pathways

POSTER 509

Islam Oguz Tuncay[1], Eun Kyoung Lee[1,2], Yoonsuh Lee[1], June-Young Koh[1], Wonchul Lee[1], Andrew Quitadamo[3], Sangmoon Lee[1], Kamran Shazand[3]

1 Inocras Inc., San Diego, CA
2 Department of Ophthalmology, Seoul National University College of Medicine, Seoul National University Hospital, Seoul, Korea
3 Shriners Hospitals for Children Headquarters, Tampa, FL

Adolescent idiopathic scoliosis (AIS) is the most common nondegenerative spinal abnormality with a prevalence of 1–4%. It is diagnosed when a three-dimensional spinal deformity measured by the Cobb angle exceeding 10° of magnitude in the coronal plane. Research indicates a strong correlation between genetics and AIS, with sibling recurrence risk of ~17.7%, and heritability estimates of ~87.5%. However, the rarity of shared causative genes among patients, and the difficulty of replication between studies and cohorts suggest that AIS is a highly complex polygenic disease that results from the interaction of multiple gene loci and the environment. In this study, we utilized whole-genome sequencing (WGS) to comprehensively explore the genetic landscape of 119 AIS patients from 103 families. We implemented an automated WGS analysis pipeline powered by RareVisionTM consisting of automated algorithms and manual curation. We identified clinically relevant candidate variants in 20/119 (16.8%) patients from 18/103 (17.5%) families, and potentially relevant strong or moderate candidate variants in another 73/119 (61.3%) patients from 61/103 (59.2%) families. Candidate variants included coding and noncoding point mutations, along with structural variants and large indels, showing the utility of WGS. Candidate genes included previously AIS-associated genes (e.g. CHD7, COL11A1/2, FBN1/2, HSPG2, KIF7), as well as genes associated with other musculoskeletal and developmental syndromes where scoliosis is a known symptom (e.g. RYR1, GJB2, MYH2, MYH7). Association analysis showed 4 known AIS SNPs (rs12946942, rs10756785, rs3904778, rs7633294) also correlated with AIS ($P<0.005$) in European and/or Admixed American individuals in our cohort. Finally, through gene set enrichment analysis of 114 candidate genes, we were able to identify 3 gene clusters involved in skeletal muscle contraction, extracellular matrix composition, and gene expression regulation. In summary, through scalable WGS-based familial testing we were able to (1) find clinically relevant genetic variations in the majority of our patients and (2) create a large dataset that allowed us to identify biological pathways relevant to AIS etiopathogenesis. Our study shows the strength of WGS in understanding rare complex traits.

Whole-genome sequencing reveals the homologous recombination deficiency in human cancer

POSTER 510

Young Seok Ju[1], Sangmoon Lee[1], Joonah Lim[1], Ryul Kim[1]

1 Inocras Inc., San Diego, CA

Breast cancer is the leading cause of cancer death in female worldwide. A substantial proportion of breast cancers are known to be deficient in the process of homologous recombination pathway, often sensitive to PARP inhibitors. There is emerging evidence that a significant proportion of homologous recombination deficiency (HRD) breast cancer patients are not BRCA1/BRCA2 carriers but sporadic, and these cancers may benefit from the targeted therapy. To understand the landscape of HRD breast cancers, we whole-genome sequenced ~1,400 breast cancers and their matched germline. Of these, ~20% cases showed clear mutational signatures attributable to HRD. To reveal their origins, we explored germline, somatic and epigenetic variants in the HR pathway genes using whole-genome and whole-genome methylation sequencing. Further, we observed similar, but slightly different mutational signatures according to the causal variant of the HRD. In this session, we will share the origins and resultant mutational signatures of HRD breast cancers.

Leveraging Biobanks for Phenotype-driven Variant Interpretation

POSTER 511

Cecile Avery[1], Leigh Anne Tang[1], Ida Aka[1], Jing He[1], Rory Tinker[1], Rizwan Hamid[2], Joy Cogan[2], John A. Philips[2], Douglas Ruderfer[1], Lisa Bastarache[1]

1 Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN
2 Department of Pediatrics, Division of Medical Genetics and Genomic Medicine, Vanderbilt University Medical Center, Nashville, TN

Introduction: Despite significant advancements in the diagnosis of genetic disease through clinical sequencing, a challenge remains in the abundance of candidate variants considered during diagnosis. Biobanks linked to electronic health records (EHR) provide an opportunity to generate evidence of co-segregation of a potentially pathogenic variant and phenotype. We developed a customizable pipeline to conduct case/control studies that aid in rare variant interpretation.

Methods: We generated phenotype risk scores (PheRS) based on clinical symptoms of disease as defined in the Online Mendelian disease in Man (OMIM). Human phenotype ontology terms from OMIM were mapped to phecodes from the EHR to generate a PheRS, which represents a weighted sum of OMIM symptoms present for a given disease at the individual level. PheRS were averaged across all individuals with the same variant and compared to patients without that variant. We evaluated the performance of PheRS using area under the receiver operator cuve (AUROC) and the area under the precision recall curve (AUPRC).

Results: We selected ClinVar benign and pathogenic variants (n=4009) in 14 genes implicated in dominant, multi-system disorders to simulate the use of PheRS on phenotypically heterogenous diseases. Using the eMERGESeq cohort (n = 25000) PheRS achieved an AUROC of 0.79 and AUPRC of 0.48. PheRS scores had similar positive and negative predictive values to CADD at their respective ROC optimal thresholds (PheRS PPV=0.35, NPV = .98, CADD PPV = .40, NPV = 1). Additionally, the scores were not correlated (Spearman's $\rho = 0.02$, $p = 0.17$), indicating that PheRS contributes independent information to variant interpretation.

Discussion: This study suggests that EHR-linked biobanks can be used to generate in silico predictions of variant pathogenicity. Notably, PheRS can be defined using existing phenotype profiles in resources like OMIM or a personalized phenotype and has been applied to challenging cases in the undiagnosed diseases network. This approach can also generate statistical evidence to support VUS resolution, ultimately improving the interpretability of clinical sequencing for all patients.

Beyond the gold standard: clinical WGS CNVs confirmation with long-read sequencing

POSTER 512

Yueyao Gao[1], James Emery[1], Jonna Grimsby[1], Michelle Cipicchio[1], Sophie Low[1], Kiran V. Garimella[1], Jonathan T. Smith[1], Katie Larkin[1], Sean Hofherr[1], Mark Fleharty[1], Niall Lennon[1]

1 Broad Institute of MIT and Harvard, Cambridge, MA

Accurate detection of copy number variants (CNVs) from NGS is crucial for enhancing clinical utility and diagnostic yield. While 30x short-read whole genome sequencing (WGS) is widely used for CNV detection, regulatory bodies and industry best practices still require additional confirmatory validation due to the high risk of false results. Traditional approaches such as karyotyping and array-based techniques are considered gold standards for CNV confirmation but are often time-consuming, labor-intensive, and limited by resolution or probe availability, leading to misreported clinical variants. This highlights the need for complementary technologies for CNV validation.

To address these challenges, we investigated long-read sequencing technologies to validate CNVs detected in short-read WGS, given its advantage of resolving complex structure variants (SV). In the first phase of our study, we analyzed a Coriell cell-line sample NA13590 containing four array-reported CNVs, ranging from 100 kb to 3 Mb, previously confirmed by karyotyping and in situ hybridization. We evaluated 30x Illumina (ILMN) clinical WGS (cWGS), 18x PacBio Revio WGS, 19x Oxford Nanopore Technologies (ONT) WGS, and an equivalent sample utilizing ONT Adaptive Sampling (ONT AS) that yielded 82x average per-target coverage, to determine whether these technologies could corroborate the array-confirmed CNV events. Our findings demonstrated that 30x ILMN cWGS successfully detected all four CNVs, validating the robustness of our cWGS approach. However, neither long-read WGS technologies consistently confirmed every CNV, whereas the ONT AS provided better recovery due to its higher depth

In the second phase, we examined a healthy donor saliva-derived sample using 30x ILMN cWGS, 8x Revio WGS, and high coverage targeted ONT AS to simulate a real clinical sample. We specifically assessed whether long reads could confirm low-confidence CNV calls from ILMN cWGS data. Our results indicate that Revio WGS and ONT AS effectively supported some of the low-confidence CNV calls, further demonstrating their utility in CNV orthogonal validation.

Our work demonstrates the potential for incorporating long-read sequencing for CNV validation. Further studies will be needed to fully understand the cost, performance, and operational trade-offs that need to be navigated in the selection of a method to support clinical production for real cases.

High-risk pregnancies enriched for genetic disorders in BeginNGS genome-based newborn screening pilot clinical trial

POSTER 601

Rebecca Reimers[1,2,9], Meredith Wright,[1,2] Lauren Olsen,[1,2] Brandan Schultz,[1,2] Liana Protopsaltis,[1,2] Kristen Wigby,[1] Dan Averbuj,[1,2] Eric Blincow,[1,2] Jeanne Carroll,[2,3] Sara Caylor,[1,2] Thomas Defay[4], Katarzyna Ellsworth,[1,2] Annette Feigenbaum,[1,2,3] Erwin Frise,[5] Mia Gover,[1] Lucia Guidugli,[1,2] Christian Hansen,[1,2] Hugh Kwon,[1,2] Lakshminarasimha Madhavrao,[1,2] Chris M. Kunard,6 Jeremy Leipzig,[7] Yupu Liang,[4] Rebecca Mardach,[1,2,3] William R. Mowrey,[4] Hung Nguyen,[1,2] Anna-Kasia Niemi,[1,2,3] Danny Oh,[1,2] Muhammed Saad,[1,2] Gunter Scharer,[1] Jennifer Schleit,[1] Shyamal S. Mehtalia,[6] Laurie D. Smith,[1] Lucitia Van Der Kraan,[1,2] Mary J. Willis,[1] Mark Yandell[8], Stephen F. Kingsmore*[1,2]

1 Rady Children's Institute for Genomic Medicine, San Diego, CA
2 Rady Children's Hospital, San Diego, CA
3 Department of Pediatrics, University of California San Diego, San Diego, CA
4 Alexion, Astra Zeneca Rare Disease, Boston, MA
5 Fabric Genomics, Inc., Oakland, CA
6 Illumina, Inc., San Diego, CA
7 TileDB Inc., Cambridge, MA
8 Department of Human Genetics, University of Utah, Salt Lake City, UT
9 Scripps Research Translational Institute, La Jolla, CA

The clinical utility of genome sequencing (GS) for screening and diagnosis in the perinatal-neonatal care transition is still being defined. There are ongoing large clinical trials for GS in prenatal diagnosis of structural anomalies and in newborns for screening (NBS) for pediatric-onset disorders. We have separately presented findings from the pilot BeginNGS study of GS-based NBS which screens for 53,575 variants in 342 genes for 412 childhood genetic diseases where effective therapeutic interventions are available. We enrolled 120 neonates in a level IV regional neonatal intensive care unit (NICU) at Rady Children's Hospital in San Diego, California. Enrollees received rapid diagnostic genome sequencing (RDGS), the BeginNGS test, and routine clinical NBS. Neonates where clinical RDGS was under clinical consideration, where clinical RDGS was ordered, or age greater than 10 days of life were excluded. Here we report the predictive value of abnormal features of pregnancy for genetic diagnosis in newborns. Pregnancy features were abstracted from the newborn's medical record.

(continue to page 109)

(continued from page 108)

Of 24 neonates with pathogenic or likely pathogenic (P/LP) RDGS results, 14 had a composite obstetric comorbidity that included abnormal antepartum genetic screening, assisted reproductive technology, multiple gestation, structural ultrasound findings, prior intrauterine fetal demise or neonatal death, recurrent pregnancy loss (2+), or amniotic fluid abnormalities.

The relative risk of a diagnostic neonatal RDGS was 3.74 (95% CI 1.81-7.74, P=0.0004) in those with this composite pregnancy morbidity. The results in these infants included P/LP single gene variants in CSNK2A1, PAH, PALB2, PIK3R2, PDX1, PROKR2, SCN4A, TSHR, a 9p24.1 deletion, a 22q11.2 duplication, and Trisomy 21. There were no statistically significant differences in gestational age or birthweight between the diagnostic (P/LP) and nondiagnostic results.

Of the RDGS diagnostic findings, only two of the genes (PAH, CYP21A) are on ACMG Tier 3 carrier screening panels (carrier frequency greater or equal to 1/200). None of the detected disorders are clinically available on single gene noninvasive prenatal offerings (Vistara™, Unity Fetal Risk Screen™).

The results of this pilot study suggest that diagnostic RDGS results are significantly increased in those with pregnancy morbidity. Larger prospective studies are needed to validate utility of composite obstetric comorbidity as an indication for neonatal RDGS.

Clinical calibration of a regional missense intolerance metric quantified from 125,748 exomes

POSTER 602

Lily Wang[1,2,3], Katherine Chao[1,2], Ruchit Panchal[1,2], Haneen Abderrazzaq[4], Calwing Liao[1,2,3], Robert Ye[1,2,3], Genome Aggregation Database Consortium, Grace Tiao[1,2], Heidi Rehm[1,2,3], Mike Talkowski[1,2,3], Daniel MacArthur1,3,5,6, Anne O'Donnell-Luria[1,2,3,7], Konrad Karczewski[1,2], Predrag Radivojac[4], Mark Daly[1,2,3], Kaitlin Samocha[1,2]

1 Broad Institute of MIT and Harvard, Cambridge, MA
2 Massachusetts General Hospital, Boston, MA
3 Harvard Medical School, Boston, MA
4 Khoury College of Computer Sciences, Northeastern University, Boston, MA
5 Garvan Institute of Medical Research, and UNSW Sydney, Sydney, Australia
6 Murdoch Children's Research Institute, Melbourne, Australia
7 Boston Children's Hospital, Boston, MA

Missense variants comprise the majority of coding variants in a genome and are known to be major drivers of human disease yet are notoriously difficult to interpret. Building precise tools for missense interpretation is thus crucial to understanding this variant class and improving the diagnostic rate for individuals with suspected genetic disease. Here, we quantify missense intolerance at the sub-genic level to inform potential variant consequences, leveraging 125,748 exomes in gnomAD v2.1 to search 18,629 canonical protein-coding transcripts for variability in missense constraint (number of observed rare population missense variants compared to neutral expectation, or OE).

Missense-constrained regions are enriched for ClinVar pathogenic/likely pathogenic (P/LP) missense variants. In autosomal dominant disease genes, P/LP variants tend to localize to regions that are more missense-constrained relative to the overall transcript (Wilcoxon p = 3.5x10-10), while the opposite is true for B/LB variants (Wilcoxon p < 10-18). We thus calibrated our metric for use in ACMG/AMP clinical variant classification, applying an established framework (Pejaver et al. 2022). Missense OE ⩽ 0.21 met moderate levels and OE ⩽ 0.37 met supporting levels of evidence for pathogenicity under the hotspot/functional domain (PM1) criterion.

(continue to page 111)

(continued from page 110)

While no global threshold met any evidence levels supporting in silico benignity (BP4), separate calibration specifically in transcripts with missense constraint variability demonstrated that missense OE ⩾ 1.56 met moderate evidence for BP4 and OE ⩾ 0.97 met supporting evidence, indicating that in transcripts where our approach is powered to characterize regional constraint, missense OE close to one may be suggestive of benignity. Finally, we incorporated our regional constraint metric into a missense deleteriousness predictor (MPC) and demonstrate similar performance to deep-learning based models like AlphaMissense in prioritization of de novo missense variation in early developmental phenotypes like autism and intellectual disability. We have publicly released these metrics for use in gene discovery efforts and clinical variant interpretation settings, anticipating refined resolution as datasets grow in size and ancestral diversity and with the incorporation of complementary structural or functional data.

Optimized chemistry allows for improved sensitivity for measurable residual disease (MRD) detection below 0.1% in acute myeloid leukemia

POSTER 603

Paula Kalavakur[1], Michael Washburn[1], Laura Johnson[1], Allie Hadjis[1]

1 IDT Archer, Boulder, CO

Acute Myeloid Leukemia is an aggressive and rapidly progressing cancer with an estimated 40-50% recurrence rate following initial treatment [1]. Identifying variants at allele frequencies below 0.1% enables early predictive detection. Measurable Residual Disease (MRD) describes a small number of cancer cells that remain after treatment and corresponds with higher incidence of relapse [2]. Additionally, MRD negativity is associated with increased long term survival outcomes [3]. Next Generation Sequencing is a reliable method of detection that allows for a high degree of sensitivity to monitor measurable residual disease persistence and detect the presence of recurrent mutations at low frequencies.

IDT's Archer VARIANTPlex™ assays, combined with Archer anchored multiplex PCR (AMP™) chemistry and Archer analysis, offer a reproducible high-throughput method to test for multiple variants simultaneously. Here we present updated Archer chemistry that demonstrates over 2 times lower error rates and improved coverage over hotspot regions when compared with standard chemistry. Seraseq myeloid cell line containing relevant AML mutations such as NPM1, was diluted into background NA12877 to test for lower limits of detection below 0.1%. Over 2X lower error rates were observed, allowing for increased sensitivity in variant calling below 0.1% for select variants.

References:

[1] Thol F, Ganser A. Treatment of Relapsed Acute Myeloid Leukemia. Curr Treat Options Oncol. 2020 Jun 29;21(8):66. doi: 10.1007/s11864-020-00765-5. PMID: 32601974; PMCID: PMC7324428.

(continue to page 113)

(continued from page 112)

[2] Li W. Measurable Residual Disease Testing in Acute Leukemia: Technology and Clinical Significance. In: Li W, editor. Leukemia [Internet]. Brisbane (AU): Exon Publications; 2022 Oct 16. Chapter 5. Available from: https://www.ncbi.nlm.nih.gov/books/NBK586210/ doi: 10.36255/exon-publications-leukemia-measurable-residual-disease

[3] Short NJ, Zhou S, Fu C, Berry DA, Walter RB, Freeman SD, Hourigan CS, Huang X, Nogueras Gonzalez G, Hwang H, Qi X, Kantarjian H, Ravandi F. Association of Measurable Residual Disease With Survival Outcomes in Patients With Acute Myeloid Leukemia: A Systematic Review and Meta-analysis. JAMA Oncol. 2020 Dec 1;6(12):1890-1899. doi: 10.1001/jamaoncol.2020.4600. PMID: 33030517; PMCID: PMC7545346.

Linking gene expression to longitudinal aging and metabolic phenotypes through archival blood single-nuclei sequencing

POSTER 604

Mukundh Murthy[1], Ajay Nadig[1], Nicolas Lapique[1], Abdulraouf Abdulraouf[1], Naeem Nadaf[1], Benjamin S. Williams[2], Avshalom Caspi[2], Terrie E. Moffitt[2], Evan Macosko[1,3]

1 Broad Institute of MIT and Harvard, Cambridge, MA
2 Department of Psychology and Neuroscience, Duke University, Durham, NC
3 Department of Psychiatry, Massachusetts General Hospital, Boston, MA

Longitudinal cohort studies have routinely collected frozen whole blood from subjects, offering an opportunity to track cellular and molecular changes in people as they age. Recently developed single-cell RNA-seq (scRNA-seq) methods have successfully studied cell-type-specific regulatory patterns in blood, but require special cryoprotectants to freeze blood and ensure intact cells upon thawing. We developed Blood-Seq, a protocol that isolates intact cellular nuclei from archival, frozen whole-blood samples. Comparison to published PBMC datasets revealed that Blood-Seq sampled equivalent numbers of genes per nucleus (median 1576 genes) and resolved cellular states with similar resolution.

We applied Blood-seq to the Dunedin cohort, a longitudinal study from New Zealand that began in 1972, which has collected clinical phenotypic data from participants at regular intervals since birth. We sequenced snRNA-seq profiles from 367166 frozen PBMC nuclei, across 483 study participants, collected in 1998 when participants were 26. To identify expression signatures associated with particular clinical traits, we employed differential expression and TRanscriptome-wide Analysis of Differential Expression (TRADE), a recently described statistical framework for estimating perturbational effects on the entire transcriptome. In CD4 T cells, we detected statistically significant gene expression effects due to differences in patient maximum oxygen consumption during exercise, C-reactive protein, cortisol, glycated hemoglobin, and active smoking status. Examining trait-perturbed genes revealed enrichment for known biomarkers, validating that Blood-Seq recovers meaningful expression patterns from blood frozen 26 years ago.

(continue to page 115)

(continued from page 114)

Study participants' Pace of Aging (PoA) was tracked using a composite measure from 19 aging biomarkers collected between ages 26 and 45. We investigated if significant gene expression effects at age 26 predicted PoA differences, and found that genes associated with cortisol response–through regulation by the glucocorticoid receptor NR3C1–were increased in participants with a faster PoA, exemplified by elevated FKBP5 expression. The PoA signature was also enriched for genes responsive to deletion of steroid hormone synthesis regulators, reinforcing the prominence of stress-related pathways.

Ultimately, our work unlocks a large array of archival samples for deeper molecular phenotyping, with insights into the molecular processes mediating aging.

Functional impact of 2.7M structural variants across global populations

POSTER 605

Xuefang Zhao[1,2,3], Ryan L. Collins[4], Arthur S. Lee[1,2,3], Mark Walker[1,3], Emma Pierce-Hoffman[1,3], Alex Yenkin[1,2,3], Nehir Kurtas[1,2,3], Jack M. Fu[1,2,3], Philip W. Darnowsky[3], Katherine Chao3, gnomAD production team, gnomAD consortium, Stacey Gabriel[3], Eric Banks[3], Anne O'Donnell-Luria[3,6], Daniel G MacArthur[3], Heidi L. Rehm[3], Benjamin M Neale[2,3,5], Mark J Daly[2,3,5], Kaitlin E. Samocha[1,2,3], Konrad J. Karczewski[1,2,3], Harrison Brand[1,2,3], Michael E. Talkowski[1,2,3W]

1 Center for Genomic Medicine and Department of Neurology, Massachusetts General Hospital, Boston, MA
2 Department of Neurology, Harvard Medical School, Boston, MA
3 Program in Medical and Population Genetics and Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA
4 Division of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA
5 Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA
6 Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA

Structural variants (SVs) involve large (>50 bp) rearrangement of the human genome and contribute to numerous diseases and traits. The complexity of discovering, classifying, and predicting functions of SVs from short-read genome sequencing (srGS) has largely precluded the development of SV reference maps at a scale and diversity comparable to those routinely employed for clinical screening of point mutations. This dearth of openly accessible SV reference resources leaves a significant void in the comprehensive classification of clinically relevant variation and/or trait association evidence in those individuals harboring SVs that affect coding and regulatory sequences. Here, we present an atlas of ~2.7M SVs uniformly processed from srGS of 63,046 individuals in the Genome Aggregation Database (gnomAD) and 97,940 individuals in All of Us Research Program (AoU). We benchmarked these data using long-read genome sequencing across 1,027 matched samples and microarray from 12,141 matched samples, achieving >96% precision from each. Per individual genome, we discovered an average of 11K SVs and found rare SVs (allele frequencies < 1%) causing loss of function (LoF) of 8 genes. We observed significant correlations between frequencies of rare LoF SVs and previously derived constraint metrics (LOEUF; r = 0.97), indicating strong selection against SVs that alter constrained genes.

(continue to page 117)

(continued from page 116)

Taking advantage of the power of these cohorts, we sought to define noncoding genomic regions that may be intolerant to SVs. We developed a hypothesis-free noncoding combinatorial annotation selection (NCAS) test by exhaustively annotating SVs across 233 noncoding regulatory elements leading to 472,824 unique combinations. Overall, we discovered 855 combinations that achieved Bonferroni corrected significance in gnomAD with 551 showing replication in AoU. The most significant driver of our noncoding signals are ultra conserved elements, which showed comparable depletion of mutation rates to LoF SVs.
Overall, these findings could have a substantial impact on variant interpretation and genetic diagnosis in the 98% of the genome that is noncoding, the sequences that the clinical fields have struggled to interpret. Importantly, we have released all data in this study (gnomAD: https://gnomad.broadinstitute.org/, AoU: https://gnomad.broadinstitute.org/), allowing all researchers to freely access these resources for biomedical research and clinical variant interpretation.

	Last Name	First Name	Organization / Company	Page(s)
G	Gao	Shanshan	University of Colorado Anschutz Medical Campus	90
	Gao	Yueyao	Broad Institute of MIT and Harvard	107
	Garcia	Brandon	Baylor College of Medicine	83
	Ge	Tian	Massachusetts General Hospital	82
	Goar	Wesley	Nationwide Children's Hospital	66
	Gogate	Nikhita	Baylor College of Medicine	53, 85
	Guettouche	Toumy	Mercy BioAnalytics, Inc.	19, 20
	Gustafson	Anxhela	The Shriners Children's Genomics Institute	34
H	Hage	Rosemary	Centers for Disease Control and Prevention	28, 31
	Hiatt	Laurel	University of Utah	91
	Hite	Dustin	Watchmaker Genomics	68
	Hou	Ying-Chen Claire	Nationwide Children's Hospital	25
J	Jayasinghe	Reyka	Washington University School of Medicine	36, 37, 44
	Ju	Young Seok	Inocras Inc.	105

	Last Name	First Name	Organization / Company	Page(s)
N	Noronha	Katelyn	AtlasXomics Inc	85
O	O'Donnell-Luria	Anne	Boston Children's Hospital	100, 110
P	Page	Nick	Kavli Institute for Fundamental Neuroscience	55
	Pater	Megan	Saint Louis University School of Medicine	87
	Pennacchio	Alex	Lawrence Berkeley National Lab	78, 95, 100
	Polfus	Linda	Ambry Genetics	65
R	Reimers	Rebecca	Rady Children's Institute for Genomic Medicine	76, 108
	Relova	Rona Margaret	United States Department of Veterans Affairs (VA)	47
	Reva	Boris	Icahn School of Medicine at Mount Sinai	59, 60
S	Salman	Areesha	The Broad Institute of MIT and Harvard	27
	Santos	Jerome	Alida Biosciences	39
	Sederman	Casey	University of Utah	29
	Singh	Mugdha	Boston Children's	52
	Slaven	Neil	HospitalLawrence Berkeley National Lab	78

	Last Name	First Name	Organization / Company	Page(s)
S	Stamper	Ericca	Dovetail Genomics	98, 99
	Stenton	Sarah	The Broad Institute of MIT and Harvard	57, 58, 80
	Stephenson	Kristi	Stephenson	21, 22
T	Takemon	Yuka	University of British Columbia	51
	Tawa	Gregory	National Institutes of Health (NIH)	61
	Tee	Xin Ray	GenScript USA Inc.	38
	Teng	Shaolei	Howard University	48
	Tran	Thinh	Natera, Inc.	84
	Tuncay	Oguz	Inocras Inc.	104
V	Verma	Rashi	Morehouse School of	93, 94
W	Wang	Lily	Broad Institute of MIT and Harvard	110, 111
	Wright	Meredith	Rady Children's Institute for Genomic Medicine	76, 77, 108
	Wu	Haojia	Washington University in St. Louis	30
Y	Yang	Jiyuan	St. Jude Children's Research	88, 89
Z	Zhao	Xuefang	Harvard Medical School	100, 116, 117
	Zhou	Wei	The Rockefeller University	32, 33, 67

Unlock the *potential* of precision health data to drive scientific breakthroughs

DNAnexus® is helping hundreds of biomedical organizations accelerate scientific discovery. With DNAnexus, precision health innovators are able to manage, analyze, and collaborate on multimodal data to harness its full potential and unlock insights that improve patient care.

JOIN US FOR SESSION VI:	Fireside Chat – A Patient's Perspective • 9/6 at 12:15 pm in Adams Ballroom

DNAnexus®

dnanexus.com

AGBT2025™
AGRICULTURE

MARCH 31 – APRIL 2, 2025

CARIBE ROYALE HOTEL | ORLANDO, FL

AGBT 2025™

PRECISION HEALTH

LOEWS CORONADO
BAY, SAN DIEGO, CA

AGBT PH RETURNS TO CORONADO ISLAND IN 2025

SAVE THE DATE | SEPTEMBER 8 – 10

LEARN MORE AT AGBT.ORG