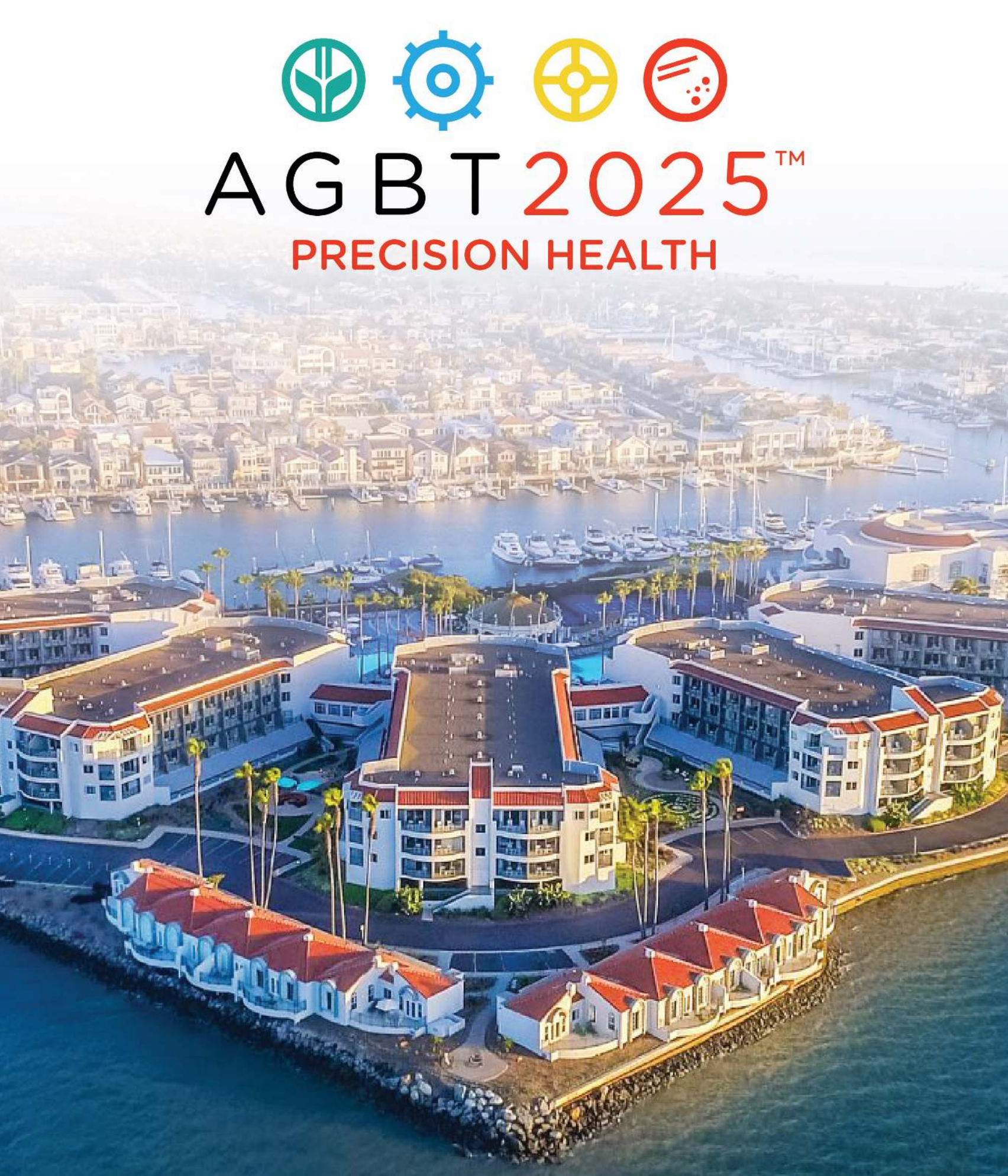
AGBT2025™

PRECISION HEALTH



September 8–10, 2025 | Loews Coronado Bay Resort | San Diego, CA

HITACHI

Nabsys

Nabsys, AGBT Precision Health,
and the GREGoR Consortium invite you to join us at

Sips & Science: A Sunset Cruise

With dinner, drinks, and coastal sunset views

Wednesday, September 10 | 6 – 9 PM | Departing from Loews dock

Meet with our team during AGBT PH

Learn more about our unprecedented genome mapping technology

Justin Bingham
Chief Commercial Officer

Edd Lee
Global Vice President, Sales and Marketing

Jason Hughes, PhD
Vice President, Computational Biology

John Thompson
Principal Application Scientist

Elise Benson
Director, Marketing

Barrett Bready, MD
Founder/CEO

Nabsys Hospitality Suite

Join us during the Welcome Cocktail Hour, dessert, or Sponsor Social

Sovereign Suite

Monday, 4 – 5 PM
Tuesday, 1 – 2 PM & 8 - 11 PM

Electronic Genome Mapping for Whole Genome Structural Variant Discovery and Confirmation

John Thompson
Principal Application Scientist

Main Stage
Tuesday, 2:20 – 2:30 PM

Schedule a meeting or request more information

nabsys.com

AGBT2025™
PRECISION HEALTH

Welcome to the 2025 Advances in Genome Biology and Technology (AGBT) Precision Health Meeting!

You are the leaders shaping the future of precision medicine. For 10 years, the AGBT Precision Health Meeting has brought together scientists, clinicians and innovators driving genomic discoveries from the lab to the clinic—and we are honored to have you with us for this milestone year.

This meeting unites geneticists, principal investigators, medical professionals, healthcare innovators, and biotech executives from around the world. Together, you help move breakthrough genomic technologies into real-world patient care.

This meeting is organized by The Genome Partnership, a not-for-profit dedicated to advancing research, promoting education, and expanding commerce in genome science and technology. Learn more at agbt.org.

On behalf of The Genome Partnership and the AGBT PH25 Scientific Organizing Committee, we're honored to have you here.

Wendy Vlieks

Wendy Vlieks
President, The Genome Partnership
Host of AGBT Meetings

Organizing Committee

We thank our Meeting Co-Chairs Mike Talkowski, Eric Green, and Wendy Chung and the Scientific Organizing Committee for their time and effort in developing the AGBT Precision Health 2025 program.

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
Former-National Human Genome Research Institute

Sue Hill
NHS, England

Gail Jarvik
University of Washington

Stephen Kingsmore
Rady Children's Institute for Genomic Medicine

Howard Mcleod
Utah Tech University

Stephen Montgomery
Stanford 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

AGBT2025™
PRECISION HEALTH

THANK YOU FOR THE GENEROUS SUPPORT FROM OUR 2025 SPONSORS:

MAJOR SPONSORS

Nabsys

SeqOne
FROM GENOME TO PERSONALIZED MEDICINE

CONTRIBUTING SPONSORS

Agilent

FABRIC GENOMICS

NEW ENGLAND Biolabs®

General Information

Hospitality Desk Hours (Atrium)

- Monday, September 8: 11:00 am – 7:30 pm
- Tuesday, September 9: 7:30 am – 7:30 pm
- Wednesday, September 10: 7:30 am – 5:00 pm

Name Badges

Your badge is required for security and networking. Wearing it clearly not only keeps the meeting safe but also helps you make valuable connections.

- Please wear your badge on the provided lanyard with your name facing outward.
- Keep it with you at all times; it's required for entry to all sessions and events.
- Badges are non-transferable.

Transportation

The Loews Coronado Bay Resort is approximately 12 miles from San Diego International Airport (SAN). The drive typically takes about 20 minutes, depending on traffic. Transportation options include Uber, Lyft, and taxi service.

Social Events

- **Welcome Cocktail Hour:** Monday, Sept. 8, 4–5 p.m. – Constellation Foyer
- **Welcome Reception and Dinner:** Monday, Sept. 8, 7:10–10 p.m. – Bay Terrace
- **Poster Session, Sponsor Reception & Dinner:** Tuesday, Sept. 9, 4–8 p.m. – Constellation Ballroom & Foyer
- **Sponsor Social:** Tuesday, Sept. 9, 8–11 p.m. – Constellation Foyer
- **Women's Networking Event (RSVP Required):** Wednesday, Sept. 10, 7:30–9 a.m. – Avalon
- **Sips & Science: A Sunset Cruise with AGBT & GREGoR (RSVP Required):** Wednesday, Sept. 10, 6–9 p.m. - Yacht

Meeting Technology

Official AGBT Mobile App

Enhance your meeting experience with our comprehensive mobile app, designed to keep you connected before, during and after the meeting.

Features:

- Interactive agenda
- Speaker and abstract details
- Networking tools
- Real-time schedule updates
- Interactive venue map

Download Instructions:

- iOS: Scan the QR code or download from the App Store
- Android: Scan the QR code or download from Google Play
- Alternative: Scan the QR code at the registration desk or ask for assistance

Social Media Engagement

Stay connected and share your experience at AGBT Precision Health! Join the conversation using **#AGBTPH25**.

Wi-Fi Access

- Network: **AGBT Precision Health**
- Password: **AGBTPH25**

Capture the moment, connect with colleagues, and celebrate scientific advancement!

SeqOne

ISO 13485 · ISO 27001 · GDPR · HIPAA

Smarter Genomics.

Faster Insights. Greater Confidence.

For somatic and germline analysis

Wet lab and sequencer agnostic

Clinically validated, integrating top-tier databases

Premium workflow automation

Agenda

Monday, September 8

11:00 am – 7:30 pm	**Hospitality Desk** **Location:** Atrium
1:00 pm – 4:00 pm	**Pre-Conference Workshops** "Join hands-on, expert-led sessions exploring tools, case studies, and practical approaches in precision health."
1:00 pm – 2:30 pm	**Genomic Sequencing for Rare Disease: Best Practices for Analysis and Interpretation** **Speakers:** Christian Marshall, PhD, FACMG, FCCMG - Molecular Laboratory Director, Division of Genome Diagnostics, The Hospital for Sick Children (SickKids) Vaidehi Jobanputra, PhD, FACMG - Chief Diagnostics Officer, New York Genome Center and Associate Professor, Columbia University Medical Center **Moderator:** Katarzyna Ellsworth, PhD, FACMG - Senior Director of Clinical Operations, Rady Children's Institute for Genomic Medicine **Workshop Description:** This session will explore the current state of genomic sequencing for rare disease diagnosis, including technical and analytical standards, validation strategies and best practices for interpretation and reporting. It will include a discussion of current challenges and how emerging technologies, including multiomics and long-read sequencing, will impact diagnosis. **Location:** Commodore Ballroom

2:30 pm – 4:00 pm	**Leveraging the All of Us Resources to Advance Precision Medicine** **Speakers:** Alex Bick, Division of Genetic Medicine at Vanderbilt University Medical Center Tara Dutka, NIH All of Us Research Program **Workshop 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. **Location:** Commodore Ballroom
4:00 pm – 5:00 pm	**Welcome Cocktail Hour with Sponsors, Sponsored by New England Biolabs** **Location:** Constellation Foyer
5:00 pm – 5:10 pm	**Opening Remarks** (AGBT PH Co-Chairs) **Location:** Commodore Ballroom Wendy Chung, Boston Children's Hospital Eric Green, Former Director, NHGRI Michael Talkowski, Massachusetts General Hospital, Broad Institute
Session I:	**Genomic Medicine Today and Tomorrow** "This session sets the stage by reflecting on the evolution of genomic medicine and its influence on healthcare today, highlighting key milestones and future directions." **Session Chairs:** Michael Talkowski & Eric Green
5:10 pm – 5:40 pm	**Eric Green, Former Director, NHGRI** "10 Years of the AGBT Precision Health Meeting: Witnessing the Realization of Genomic Medicine "

5:40 pm – 6:10 pm	**Mary-Claire King, University of Washington** "(Almost) All Diseases Are Genetically Rare"
6:10 pm – 6:40 pm	**Benjamin Kleinstiver, Mass General Hospital & Harvard Medical School** "Engineering CRISPR Tools for Bespoke Gene Editing Therapies & Beyond"
6:40 pm – 6:55 pm	**Danny Miller, University of Washington (Abstract Selected Talk)** "Long-read sequencing as a single test in the clinical setting "
6:55 pm – 7:10 pm	**Harrison Brand, Mass General Brigham (Abstract Selected Talk)** "Introduction of Noninvasive Fetal Exome Sequencing into Prenatal Care"
7:10 pm – 10:00 pm	**Welcome Reception & Buffet Dinner** **Location:** Bay Terrace

Tuesday, September 9

AI in Precision Health
"We'll explore how AI shapes genomics and clinical care—its promise, limits, and where expert judgment remains essential."

6:30 am – 7:30 am	**Rise & Shine with AGBT** (optional activity)
7:30 am – 7:30 pm	**Hospitality Desk** **Location:** Atrium
7:30 am – 9:00 am	**Breakfast** **Location:** Commodore Terrace

Session II:	**AI and Precision Health: Honoring the Vision of Atul Butte** "To explore cutting-edge applications of AI in precision health, spanning genomics, clinical informatics, multi-omics integration, and ethical implementation: building on the vision and legacy of Dr. Atul Butte." **Session Chairs:** Geoff Ginsburg, NIH All of Us Research Program Stephen Kingsmore, Rady Children's Institute for Genomic Medicine **Location:** Commodore Ballroom
9:00 am – 9:05 am	**Welcome** (Geoff Ginsburg, NIH All of Us Research Program)
9:05 am – 9:30 am	**Euan Ashley, Stanford University** "The Next Frontier of AI in Precision Health"
9:30 am – 9:45 am	**Chris Longhurst, University of California San Diego** "AI for Learning Health Systems and Patient-Centered Care"
9:45 am – 10:00 am	**Lucila Ohno-Machado, Yale School of Medicine** "AI-Driven Biomedical Data Science: From Discovery to Translation"
10:00 am – 10:15 am	**Marina Sirota, Bakar Computational Health Sciences Institute (BCHSI), University of California, San Francisco** "AI, Clinical Data and Multi-Omics for Precision Health"

10:15 am – 10:30 am	**Bimal Chaudhari, Nationwide Children's Hospital (Abstract Selected Talk)** "HPO and PheCode-based machine learning models predict need for genetic testing in babies admitted to Level IV NICUs"
10:30 am – 11:00 am	**Coffee Break** **Location:** Commodore Foyer and Terrace
Session III:	**The Great Debate**
11:00 am – 11:50 am	**Will AI change rare disease diagnosis?** "This interactive panel explores whether AI can truly transform rare disease diagnosis—enhancing accuracy, speeding timelines, and improving outcomes—or if the field remains premature." **Session Chair & Moderator:** Stephen Kingsmore – President & CEO, Rady Children's Institute for Genomic Medicine **Panelists:** • Christian Marshall, The Hospital for Sick Children (SickKids) • Euan Ashley, Stanford University • Marina Sirota, Bakar Computational Health Sciences Institute (BCHSI), University of California, San Francisco **Location:** Commodore Ballroom
11:50 am – 12:05 pm	**Johnny Yu, Tahoe Therapeutics (Abstract Selected Talk)** "Tahoe-100M: A Giga-Scale Single-Cell Perturbation Atlas for Context-Dependent Gene Function and Cellular Modeling"

12:05 pm – 2:00 pm	**AGBT Lunch** Commodore Terrace
1:00 pm – 2:00 pm	**Dessert with Sponsors** **Location:** Constellation Foyer
2:10 pm – 2:30 pm	**Sponsor Talks** "Hear from industry partners advancing tools, platforms, and diagnostics supporting genomic innovation." **Session Chair:** Howard McLeod, Utah Tech University **Location:** Commodore Ballroom
2:10 pm – 2:20 pm	**Major Sponsor: Nabsys** **John Thompson, Principal Application Scientist** "Electronic Genome Mapping for Whole Genome Structural Variant Discovery and Confirmation"
2:20 pm – 2:30 pm	**Major Sponsor: SeqOne** **Michael Vishnevetsky, PhD, VP of Business Development, North America** "SeqOne DiagAI - Next Generation of AI-Powered Bioinformatics"
Session IV:	**Transformative Advances in Rare Disease Genomics** "This session highlights recent breakthroughs in identifying, characterizing, and treating rare genetic diseases, including insights from clinical trials, functional genomics, and emerging therapies." **Session Co-Chairs:** Heidi Rehm, Harvard, Broad Institute Wendy Chung, Boston Children's Hospital

2:30 pm – 2:55 pm	**Heidi Rehm, Harvard, Broad Institute** "Global Strategies to Solve Rare Disease"
2:55 pm – 3:20 pm	**Patricia Dickson, Washington University School of Medicine and Inez Oh, PhD, Senior Scientist, Institute for Informatics, Data Science, and Biostatistics, Washington University School of Medicine** "Computational and AI Approaches to Clinical Trial Data Harmonization for Rare Disease"
3:20 pm – 3:45 pm	**Jonathan Sebat, UCSD** "Linking genotype to phenotype in rare disease requires integrative analysis of pathway, cell-type and spatial gene expression"
3:45 pm – 4:00 pm	**Poster Flash Talks** **Anne O'Donnell-Luria, Broad Institute / Boston Children's Hospital** "Calibrating evidence to improve variant classification for genomic medicine" **Paul Valdmanis, University of Washington** "An enhancer deletion at APOE is protective against Alzheimer's disease but not lipid-related traits in African Americans" **Philip Boone, Massachusetts General Hospital** "Aberrant recursive splicing in a human disease locus" **Boxun Zhao, Boston Children's Hospital** "Pathogenic Transposon Insertions and Individualized Intravitreal Antisense Oligonucleotide Therapy for FLVCR1-Related Retinitis Pigmentosa" **Meredith Wright, Rady Children's Institute for Genomic Medicine** "Federated data querying implementation for improving performance and clinical validation of genomic newborn sequencing"

4:00 pm – 8:00 pm	**Poster Session, Sponsor Reception & Dinner** **Hors d'oeuvres & Posters 4:00 p.m. – 6:00 p.m.** • **Odd # Posters 4:00 p.m. – 5:00 p.m.** • **Even # Posters 5:00 p.m. – 6:00 p.m.** **Dinner & Desserts from 6:00 p.m. – 8:00 p.m.** **Location:** Constellation Ballroom & Foyer
8:00 pm – 11:00 pm	**Sponsor Social** Sponsor Promenade **Location:** Constellation Foyer
Wednesday, September 10	**Genomic Medicine in Rare Disease** "Sessions focus on breakthroughs in diagnosis, functional genomics, and treatment strategies."
7:30 am – 5:00 pm	**Hospitality Desk** **Location:** Atrium
7:30 am – 9:00 am	**Breakfast** **Location:** Commodore Terrace
7:30 am – 9:00 am	**AGBT Women's Networking Breakfast** (RSVP Required) Join us for the inaugural Women's Networking Event at Precision Health – an empowering breakfast gathering dedicated to celebrating and connecting women in science and precision health. This event offers a welcoming space for conversation, mentorship, and shared experiences. **Location:** Avalon
Session V:	**New Perspectives in Precision Health** "This session explores the role of environment, nutrition, and early-life exposures in shaping precision health strategies." Session Co-Chairs: Stephen Montgomery, Stanford University Gail Jarvik, University of Washington **Location:** Commodore Ballroom

9:00 am – 9:25 am	**Neil Hanchard, Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health** "Advancing Precision Health in the Pediatric Complex Disease Space"
9:25 am – 9:50 am	**Meghan Azad, Pediatrics and Child Health, University of Manitoba** "Human Milk for Precision Health"
9:50 am – 10:15 am	**Leanne Redman, Louisiana State University** "Can we personalize nutrition? The who, what, where and when of the NIH Nutrition for Precision Health Research Consortium"
10:15 am – 10:30 am	**Craig Smail, Children's Mercy Hospital (Abstract Selected Talk)** "Mapping Parent of Origin Methylation by Long-Read Sequencing Reveals Novel Imprinting and Insight into Pediatric Disease"
10:30 am – 10:45 am	**Shira Rockowitz, Boston Children's Hospital (Abstract Selected Talk)** Diagnostic Benefits of Oxford Nanopore Long-Read Sequencing for Short-Read Negative CRDC Rare Disease Cases"
10:45 am – 11:00 am	**Rebecca Reimers, Rady Children's Institute for Genomic Medicine (Abstract Selected Talk)** BeginNGS: Innovating on interpretation, enrollment, and results return in an ongoing single arm prospective newborn sequencing by genome sequencing trial

11:00 am – 11:30 am	**Coffee Break** **Location:** Commodore Foyer and Terrace
Session VI:	**From Sequencing to Systems: A Conversation on AI in Genomic Medicine** "Join a candid discussion on the evolution of genomic technology, the rise of AI in healthcare, and what's next for data-driven innovation in precision health."
11:30 am – 12:15 pm	**Fireside Chat with Jay Flatley – Former CEO and Chairman of Illumina; Board Member, Verily** **Moderator:** Eric Green – Former Director, NHGRI
12:15 pm – 2:00 pm	**AGBT Lunch & Workshop**
12:30 pm – 1:50 pm	**Multiomics Analysis Approaches and Harmonization with Publicly Available Data for Precision Health** **Speakers:** Jon LoTempio, University of California, Irvine Danny Miller, University of Washington Anne-O'Donnell-Luria, The Broad Institute Malene Lindholm, Stanford University David Amar, Tel Aviv University **Moderators:** Matthew Wheeler, Stanford University **Workshop description:** "The workshop will highlight data and approaches to facilitate multiomic analysis through initiatives including GREGoR Consortium and MoTrPAC that leverage multiomic data. Included will be discussion of harmonization across datasets, analysis approaches for outlier and integrative discovery, and ongoing challenges."

Wednesday, September 10 **(Afternoon)**

Session VII:

Variants to Function and Precision Medicines in Rare Disease
"Discover the latest work translating genetic variants into actionable insights and precision therapies for rare diseases, spanning functional studies, target discovery, and individualized treatment."
Session Co-Chairs:
Penelope Bonnen, Baylor College of Medicine
Len Pennacchio, Lawrence Berkeley National Labs
Location: Commodore Ballroom

2:00 pm – 2:25 pm

Stephen Montgomery, Stanford University
"Multi-omics strategies for diagnosis of rare diseases"

2:25 pm – 2:50 pm

Heather Mefford, St. Jude Children's Research Hospital
"Precision diagnosis to targeted therapy in pediatric epilepsies"

2:50 pm – 3:15 pm

Timothy Yu, Boston Children's Hospital and Harvard Medical School
"Towards Interventional Genetics"

3:15 pm – 3:30 pm

Jacob Kitzman, University of Michigan (Abstract Selected Talk)
"Systematic resolution of clinical variants in hereditary cancer syndromes using pathway-scale functional assays"

3:30 pm – 3:45 pm	**Closing Reflections** **Meeting Co-Chairs:** Wendy Chung, Boston Children's Hospital Eric Green, Former NHGRI Mike Talkowski, Broad Institute of MIT and Harvard
6:00 pm – 9:00 pm	**Sips & Science: A Sunset Cruise with AGBT & GREGoR, Sponsored by Nabsys** "A unique, crossover networking event bringing together attendees from both AGBT PH and the GREGoR Consortium for an evening of relaxed yet meaningful connection across institutions and specialties. This 3-hour private yacht cruise includes dinner, drinks, and coastal sunset views while allowing attendees to reflect on the shared mission of advancing precision health and rare disease research. RSVP Required, space is limited."

AGBT2026™

GENERAL MEETING

FEBRUARY 23 – 26, 2026

HILTON ORLANDO BONNET CREEK ORLANDO, FL

Major Sponsor Bid Packets Now Open! More information can be found here:

WWW.AGBT.ORG

Abstracts

A haplotype-resolved view of human gene regulation

POSTER 101

Mitchell R. Vollger[1,2,] Elliott G. Swanson2,3, Shane J. Neph2, Jane Ranchalis2, Andrew B. Stergachis2-4,†

1 University of Utah, Department of Human Genetics, Salt Lake City, UT
2 University of Washington School of Medicine, Division of Medical Genetics, Seattle, WA
3 University of Washington School of Medicine, Department of Genome Sciences, Seattle, WA
4 Brotman Baty Institute for Precision Medicine, Seattle, WA
*These authors contributed equally to this work.
†Corresponding author.

Diploid human cells contain two non-identical genomes, and differences in their regulation underlie human development and disease. Here, we present approaches to decipher this regulatory complexity through two novel methodologies: Fiber-seq Inferred Regulatory Elements (FIRE) and synchronized long-read multi-omic profiling.

FIRE provides a more comprehensive view of the accessible chromatin landscape across the 6 Gbp diploid human genome, overcoming previously unrecognized biases in existing regulatory element catalogs. FIRE enables detection of haplotype-selective chromatin accessibility (HSCA), exposing novel imprinted elements lacking underlying parent-of-origin CpG methylation differences and modules that permit genes to escape X chromosome inactivation. We uncover that the human leukocyte antigen (HLA) locus harbors the most HSCA in immune cells, where we resolve transcription factor (TF) binding events disrupted by disease-associated variants. Finally, we demonstrate that the regulatory landscape of a cell is littered with autosomal somatic epimutations that are propagated by clonal expansions to create mitotically stable and non-genetically deterministic chromatin alterations.

While FIRE provides regulatory maps, understanding disease mechanisms often requires simultaneously characterizing multiple molecular layers. We developed a multi-omic approach combining long-read genome, methylome, epigenome, and transcriptome sequencing. This approach enables accurate variant calling, genome assembly, and simultaneous characterization of haplotype-resolved CpG methylation, chromatin accessibility, and full-length transcripts. To demonstrate clinical utility, we applied this to an Undiagnosed Diseases Network participant with an X;13 translocation, revealing disruption of four genes (NBEA, PDK3, MAB21L1, RB1) through fusion transcript formation, enhancer adoption, transcriptional readthrough silencing, and inappropriate X-inactivation of autosomal genes.

(continue to page 24)

(continued from page 23)

Building on haplotype-phased regulatory maps, we demonstrate that non-coding mutations in a short tandem repeat underlie dominantly inherited Resistance to Thyrotropin (RTSH) in 82 affected participants via an activated thyroid-specific enhancer cluster. This leads to haplotype-specific upregulation of the bicistronic MIR7-2/MIR1179 locus and overexpression of its microRNA products in participants' thyrocytes. The resulting signaling pathway imbalance provides a working model for this RTSH cause.

Together, these advances provide unparalleled resolution of diploid genome regulation, revealing novel disease mechanisms through comprehensive characterization of chromatin accessibility, DNA methylation, and transcriptional output.

A high-sensitivity variant and allele imputation reference for human MHC

POSTER 102

Solomon Endlich[1], Solomon M. Adams1, Julianne K. David1, Liza Huijse1

1 Base5 Genomics, Inc., Palo Alto, CA

Variant and allele imputation is a crucial step in genetic association and certain clinical workflows that rely on genotype microarrays, yet reference panels often incorporate bias from the underlying genotyping method, ancestral representation, population sampling, and more. This limits their sensitivity and applicability, especially for underrepresented populations and in highly diverse and polymorphic genomic regions like the human major histocompatibility complex (MHC), which is one of the most clinically important and association-rich regions in the human genome.

We present a custom MHC-targeted reference panel (B5-MHC) that leverages GEM, a novel pangenome graph method of generating contiguous diploid assemblies with high-throughput, low-cost short DNA-seq reads. B5-MHC is built on GEM-based variant calls, human leukocyte antigen (HLA) allele annotations, and phasing for 3,014 samples from the 1000 Genomes Project. We compare its performance against the Michigan imputation server panels (HRC/HLA-TAPAS). B5-MHC achieves unprecedented HLA gene typing breadth (45 genes vs. 7 genes for HLA-TAPAS) and accuracy (99% vs 90% for HLA-TAPAS) with comparable results for samples across five diverse subpopulations, as well as unparalleled variant calling accuracy (0.977 F1 score vs 0.727 for HRC) and sensitivity (0.970 vs. 0.576 for HRC) compared to the National Institute of Standards and Technology (NIST) benchmark. Our groundbreaking performance is driven by GEM, which improves variant calling performance compared to BWA+DeepVariant (F1 scores of 0.90 vs. 0.84 relative to long-read based variant calls, and 0.98 for both methods relative to NIST) and increases copy number accuracy (97% vs. 72%) and allele calling accuracy (90% vs. 67% at 4-field/8-digit resolution) compared to T1K in the MHC.

B5-MHC makes dramatic improvements in variant-call sensitivity and the breadth and accuracy of HLA allele typing across diverse population groups from standard microarray data. Full imputation coverage allows for a critical increase in its functional value in the highly complex MHC region.

A multi-omic investigation of gene dysregulation in a complex chromothripsis-like translocation event using Oxford Nanopore sequencing

POSTER 103

Rebecca Stubbs[1], Emma Mizrahi-Powell2, Priyesh Rughani1, Carly Tyer1, Bryan Leland1, Sean McKenzie1, Scott Hickey1, Sissel Juul1, Gilad Evrony2

1 Oxford Nanopore Technologies, Alameda, CA
2 New York University Langone Health, New York, NY

Rare genetic diseases with complex genomic signatures can be challenging to diagnose using traditional methods, typically requiring multiple complementary techniques to fully resolve and understand the underlying biological mechanisms. Further, some instances of complex genomic variation may benefit from comprehensive multi-omic data, including de novo genomic assembly, transcriptomics, and chromatin profiling for events that are not resolved by standard reference-based variant calling.

Here, we performed in-depth multi-omic profiling of a case from the Undiagnosed Diseases Program (UDP) at New York University. We combined telomere-to-telomere (T2T) assembly via ultra-long, ligation, and Pore-C sequencing with direct RNA sequencing and chromatin stenciling from Oxford Nanopore to enrich the assembly with transcriptomic and epigenomic layers. This enabled complete resolution of a complex translocation between chromosomes 2 and 16 previously identified by karyotype that was not resolved by short-read whole-genome sequencing, due to the translocation creating many chromothripsis-like breakpoints spanning several megabases. The resulting de novo telomere-to-telomere (T2T) assembly provided 33/46 complete chromosomal contigs, including two contigs which confirmed the balanced t(2;16)(p;q) translocation event and fully resolved 34 inter- and intra-chromosomal translocation breakpoints.

This analysis identified 17 candidate genes affected by structural breakpoints, which were further investigated via 5mC methylation analysis. One disrupted gene, BCL11A, has been identified as a lead candidate based on the patient's phenotype. Additionally, this gene showed hypomethylation of Exon 4, about 70 kb from a translocation breakpoint, reflecting downstream epigenetic consequences of the translocation. Subsequent direct RNA sequencing and chromatin stenciling further identified disrupted regulatory elements that were not apparent from whole genome sequencing data alone.

Overall, this study highlights the effectiveness of combining de novo T2T assembly, 5mC profiling, RNA sequencing, and chromatin stenciling for challenging undiagnosed rare disease cases with complex structural variations, such as the chromothripsis-like translocation studied here. Our study highlights the power of comprehensive multi-omic profiling to provide deep insights into the functional impact of genic disruptions and revealing novel rare disease mechanisms.

A novel and universal enzymatic fragmentation solution for DNA: long reads, short reads, FFPE, and methylation

POSTER 104

Brittany S. Sexton[1], Karen McKay1, Margaret R. Heider1, Louise Williams1, Jonathan Sanford1, Ruby Moulton1, Matthew Angel1, Bradley W. Langhorst1, Pingfang Liu1, V.K. Chaithanya Ponnaluri1

1 New England Biolabs, Inc., Ipswich, MA

DNA fragmentation is needed to make high quality sequencing libraries for short (Illumina, MGI and Element) and long read (ONT and PacBio) sequencing. Current methods for DNA fragmentation include mechanical shearing and enzymatic fragmentation. Mechanical shearing requires expensive instruments and consumables, is not automation-friendly, and results in sample loss. Enzymatic fragmentation methods have the potential to overcome these issues, however they have certain limitations. Enzymatic fragmentation can cement DNA damage present in formalin fixation paraffin embedded (FFPE) samples, introduce sequencing artifacts and erase DNA modifications (e.g., 5-methylcytosine (5mC)). Additionally, most enzymatic fragmentation methods do not generate fragments compatible with long read sequencing. Despite the commercial availability of enzymatic fragmentation solutions, they are not suitable for fragmentation of FFPE DNA, DNA modifications detection, or for generating long DNA fragments.

To address these constraints, we developed a novel enzymatic fragmentation solution that is quick, robust, automation-friendly and tunable to generate a wide-range of DNA fragment sizes. For short reads, fragmentation of FFPE DNA upstream of library preparation resulted in a reduced frequency of FFPE-damage-derived mutations compared to other fragmentation methods. Methylation marks were preserved when used upstream of 5mC library preparation. Moreover, these 5mC libraries showed improvements in library yield and CpG coverage compared to libraries produced using mechanical shearing. For long read library preparation, enzymatic fragmentation generated tunable DNA fragment sizes from 2 kb to 30 kb (depending on fragmentation time) and automation friendly. These high-quality libraries had similar or improved N50 lengths compared with mechanically sheared libraries.

This is the first universal enzymatic fragmentation method that can be used for both short and long read libraries, generates higher quality FFPE DNA libraries and maintains methylation marks. This method simplifies sample processing, increases throughput, is cost effective, and streamlines library preparation workflows to enable high sequencing quality.

A novel NGS library preparation method that preserves and reveals the native duplex end information as a potential liquid biopsy biomarker

POSTER 105

Matt Hims[1], Yu Zheng1, Garry Teeters1

1 Canal Biosciences, Lowell, MA

Most next-generation sequencing (NGS) library preparation workflows rely on end repair and A-tailing (ER/AT) to prepare DNA templates for adapter ligation. However, these enzymatic steps can introduce mutational artifacts and obscure potentially valuable sequence information by modifying native DNA termini, filling, trimming and rewriting existing ends. For biologically derived DNA such as cell-free DNA (cfDNA), preserving native ends may provide an orthogonal and informative biomarker, complementary to somatic mutations and epigenetic analysis.

Here, we present a novel library preparation method that circumvents ER/AT, enabling the preservation and characterization of native duplex DNA ends. We validated the method using synthetic and genomic DNA samples with defined terminal sequences. We then applied the method to cell-free plasma DNA samples and report the genome-wide profiles of native end sequences. To our knowledge, this represents the first comprehensive view of native duplex cfDNA termini at genomic scale. This approach provides a high-fidelity and distinctive tool to explore native end signatures as novel biomarkers for liquid biopsy applications.

A pathway for broader use of individualized allele-selective antisense oligonucleotides

POSTER 106

Shahad Olsson[1], Liana Protopsaltis1, Joseph Gleeson1-3, Qing Zhang4, Nafeesa Kahn5, Ali Crawford5, Stephen Kingsmore1, Olivia Kim-McManus1,2

1 Rady Children's Hospital, Rady Children's Institute for Genomic Medicine, San Diego, CA
2 University of California San Diego, Department of Neurosciences, La Jolla, CA
3 n-Lorem Foundation, San Diego, CA
4 Illumina, Inc., Cambridgeshire, Cambridge, United Kingdom
5 Illumina, Inc., San Diego, CA

Antisense oligonucleotides (ASOs) offer a promising approach to treating genetic diseases by targeting RNA to modify protein expression. They can be engineered to target a patient's particular genetic variant, enabling the development of individualized treatments for rare genetic diseases. However, patients often harbor rare variants that would necessitate the manufacturing of numerous different ASOs to treat each unique patient. Allele-selective ASOs that specifically downregulate the allele containing the deleterious variant responsible for disease via a gain-of-function mechanism have demonstrated clinical efficacy. A notable example comes from an n-of-1 clinical trial utilizing an individualized, allele-selective ASO that resulted in substantial seizure reduction and improvement in neurodevelopmental measures in a patient with SCN2A-related disorder (SRD). This ASO targets the mRNA for degradation by binding to the site of a common single nucleotide polymorphism (SNP) located on the same allele as the disease-causing variant. Identifying patients with the appropriate diplotype amenable to this ASO treatment necessitates accurate haplotype phasing of the target SNP and disease-causing variant. However, determining phasing accurately is often challenging with standard paired-end short-read sequencing. Novel whole genome sequencing (WGS) technology that facilitates ultra-long phasing by utilizing cluster proximity information overcomes this limitation. Using this technology to confirm the phasing of the de novo disease-causing variant and the target SNP, we analyzed a cohort of SRD patients to assess the potential reach of the ASO therapy. Nineteen SRD probands identified through rapid WGS between 2018 and 2024 were investigated. Patients' ages at diagnosis ranged from 2-weeks-old to 12-years-old. The target SNP was harbored by eight (42%) patients and three (16%) were confirmed to have the correct SNP configuration amenable to ASO treatment. With the aid of WGS for diagnosis and therapy selection, this study showcases an innovative approach that broadens the use of individualized ASOs. By targeting common SNPs rather than solely disease-causing variants, this method offers potential benefits for a wider range of SRD patients and holds promise for application in other genetic diseases.

A transcriptomics-based computational drug repositioning pipeline identifies simvastatin and primaquine as novel therapeutics for endometriosis pain

POSTER 107

Xinyu Tang[1]*, Tomiko T. Oskotsky1,2*, Erin Arthurs3*, Arpita Govil3*, Ferheen Abbasi4, Arohee Bhoja1, Daniel J. Bunis1, Abby Lau3, Jakob Einhaus5, Maïgane Diop5, Juan C. Irwin4, Brice Gaudilliere5, David K. Stevenson6, Linda C. Giudice4**, Stacy L. McAllister3**, Marina Sirota1,7**

1 University of California San Francisco, Bakar Computational Health Sciences Institute, San Francisco, CA
2 University of California San Francisco, Department of Medicine, Division of Clinical Informatics and Digital Transformation, San Francisco, CA
3 Emory University, Department of Gynecology and Obstetrics, Atlanta, GA
4 University of California San Francisco, Department of Obstetrics, Gynecology and Reproductive Sciences, Center for Reproductive Sciences, San Francisco, CA
5 Stanford University, Department of Anesthesiology, Perioperative and Pain Medicine, Stanford, CA
6 Stanford University, Department of Pediatrics, Stanford, CA
7 University of California San Francisco, Department of Pediatrics, San Francisco, CA
*Co-first authors
**Co-senior authors

Endometriosis is a pervasive and life-altering condition characterized by severe dysmenorrhea, chronic pelvic pain, and infertility. Current treatments for endometriosis-associated pain often yield limited or transient relief and are frequently associated with side effects that hinder long-term use. These limitations highlight the pressing need for new, more effective therapeutic strategies. To address this need, we employed a transcriptomics-based computational drug repositioning pipeline to systematically analyze publicly available bulk gene expression data. This approach previously identified fenoprofen as a promising candidate, which we validated in a preclinical model. Building on that framework, we prioritized two additional FDA-approved drugs, simvastatin (a lipid-lowering agent) and primaquine (an antimalarial), based on strong gene expression reversal scores and well-established safety profiles.

We assessed their therapeutic efficacy using a validated rat model of endometriosis-associated pain. Behavioral responses to graded intravaginal balloon distension (0.01–0.90 mL) were measured at baseline, post-surgery, and post-treatment in animals receiving simvastatin, primaquine, or vehicle control (N=6 per group). RNA sequencing of uteri and lesion tissues was performed to evaluate treatment-associated changes in gene expression. Both simvastatin and primaquine significantly reduced vaginal hyperalgesia, a surrogate marker of endometriosis-associated pain. Transcriptomic analysis revealed a substantial reversal of disease-associated gene expression signatures in treated animals, consistent with the in silico predictions. These findings provide preclinical evidence that simvastatin and primaquine may serve as repurposed therapeutics for the treatment of endometriosis-related pain. More broadly, this study demonstrates the potential of transcriptomics-guided drug repositioning to accelerate the identification of novel treatment strategies for complex and understudied diseases such as endometriosis.

Aberrant recursive splicing in a human disease locus

POSTER 108

Philip M. Boone[1-3*], Ricardo Harripaul1,2,*, Rachita Yadav1,2,4,5, Michael Grzybowski6, Mahmoud K. Hanafy7, Amanda C. Lee7, Esther Y. Choi7, Ryan L. Collins1,2, Oksana Polesskaya8, Nina Makhortova9, Matthew O. Larson1, Hakan Kayir7, Yizhi Wang8, Rodolfo A. Avila8, Jude A. Frie7, Amr Eed7, Abdalla M. Albeely7, Sunitha Vemuri8, Samantha M Ayoub8, John M. Lemanski1,2, Daniel Ben-Isvy1,2,10, Xuefang Zhao1,2, Alba Sanchis-Juan1,2, Maris Handley11, Serkan Erdin1,2, Celine de Esch1,2, Kiana Mohajeri1,2, Clementine Chen3, Paulina Gonzalez Tovar3, Monica Salani1,2, Mariana Moyses-Oliveira1, Derek J.C. Tai1,2, Benjamin Currall1,2, Christopher McGraw5,12, Susan Slaughenhaupt1, Ryan Doan3, Dadi Gao1,2, James F. Gusella1,2,13,14, Sandra Sanchez-Roige8,15,16, Jared W. Young8, Jibran Khokhar7, Aron M. Geurts6, Abraham A. Palmer8,15,†, Michael E. Talkowski1,2,4,5†

1 Massachusetts General Hospital, Center for Genomic Medicine, Boston, MA
2 The Broad Institute of MIT and Harvard, Program in Medical and Population Genetics, Boston, MA
3 Boston Children's Hospital, Division of Genetics and Genomics, Boston, MA
4 The Broad Institute of MIT and Harvard, Stanley Center for Psychiatric Research, Boston, MA
5 Massachusetts General Hospital, Department of Neurology, Boston, MA
6 Medical College of Wisconsin, Department of Physiology, Milwaukee, WI
7 Western University, Department of Anatomy and Cell Biology, London, England
8 University of California San Diego, Department of Psychiatry, La Jolla, CA
9 Boston Children's Hospital, Human Neuron Core, Boston, MA
10 Harvard Medical School, Division of Medical Sciences, Boston, MA
11 Massachusetts General Hospital, Harvard Stem Cell Institute Flow Cytometry Core, Boston, MA
12 Boston Children's Hospital, Department of Neurology, Boston, MA
13 Massachusetts General Hospital, Molecular Neurogenetics Unit, Boston, MA
14 Harvard Medical School, Department of Genetics, Blavatnik Institute, Boston, MA
15 University of California San Diego, Institute for Genomic Medicine, La Jolla, CA
16 Vanderbilt University, Department of Medicine, Division of Genetic Medicine, Nashville, TN
*These authors contributed equally.
†Corresponding authors

Recursive splice sites are rare motifs postulated to facilitate splicing across massive introns and shape isoform diversity of long, brain-expressed genes. The necessity of this unique mechanism remains unsubstantiated, as does the role of recursive splicing (RS) in human disease.

(continue to page 32)

(continued from page 31)

From analyses of rare copy number variants (CNVs) from almost one million individuals, we previously identified large, heterozygous deletions eliminating an RS site (RS1) in the first intron of CADM2 that conferred substantial risk for attention deficit hyperactivity disorder (ADHD) and other neurobehavioral traits. CADM2 encodes a neuronally expressed cell adhesion molecule that has repeatedly been linked genetically to ADHD-related traits. To explore the molecular impact of RS ablation in CADM2, we used CRISPR to model patient deletions and to target a critical deletion region (~500 base pairs) containing RS1 in both human induced neurons (iNs) and rats. Transcriptome analyses in unedited iNs provided a catalog of CADM2 transcripts, including novel transcripts that retained RS exons. Intriguingly, ablating RS1 altered the gradient of RNA abundance across the first intron of CADM2, decreased the level of CADM2 expression, and impacted transcript usage. Decreased CADM2 expression was reflected in reduced exon usage downstream of the RS1 site and global alteration to genes involved in neuronal processes including synapse and axon development. Given the scale of our analyses and the widespread association of CADM2 with neurobehavioral traits, we sought to validate these findings using in vivo models and found that rodent models harboring Cadm2 RS1 deletions exhibited significant changes in locomotor activity and functional brain connectivity. In summary, our analyses demonstrate a functional role for RS as a noncoding regulatory mechanism in a gene associated with a spectrum of neuropsychiatric and behavioral traits.

Achieving and maintaining target immunosuppression levels by combined pharmacogenetic and torque teno viral load testing

POSTER 109

Anna Chikova[1], Yazan Alhadid1, Abhi Chadha1, Max Mamroth1, Zhiwei Wang1, Bethany Dale1, Alexandre Vlassov1*

1 One Lambda Inc., Thermo Fisher Scientific, West Hills, CA

Tacrolimus is an immunosuppressant commonly prescribed to kidney transplant patients. We propose a novel approach for further personalizing tacrolimus treatment by combining pharmacogenetic typing obtained pre-transplant with post-transplant monitoring of immune response levels using a common non-pathogenic virus as a biomarker. Tacrolimus is metabolized by the CYP3A5 enzyme, activity of which directly influences pharmacokinetics of the drug. Three CYP3A5 alleles (*3*6*7) are associated with full or partial loss of enzymatic activity, depending on zygosity. DPWG and the CPIC agree on recommendations for initial tacrolimus dosing based on CYP3A5 genotyping for these alleles. Pre-transplant genotyping for CYP3A5 may help to plan treatment more efficiently and achieve target concentrations of tacrolimus in the patient's blood more quickly. We developed a novel qPCR assay for genotyping *3*6*7 alleles in CYP3A5 in a single 6-plex reaction. An analytical performance study using well-characterized DNA and primary human samples including whole blood and buccal swabs demonstrated the accuracy, precision, and robustness of the assay in accordance with current CLSI standards for molecular typing. Additional biomarkers enabling monitoring variations in individual response to treatment will be beneficial in long term patient management. One such biomarker is the human Alpha Torquetenovirus (TTV), a non-pathogenic virus that is highly prevalent in the human population. Multiple published clinical studies demonstrated significant associations between TTV titers and individual immunosuppression levels. Monitoring TTV levels post-transplant may therefore offer a way to further inform patient treatment decisions. Our TTV quantification test uses multiplex qPCR measuring the 27 most common groups of TTV with high sensitivity, specificity, accuracy and precision, demonstrated by analytical verification. Testing the assay against high titer (up to 5*106 copies per reaction) of closely related beta and gamma Anelloviruses demonstrated lack of detectable cross-reactivity and high specificity for TTV. Linearity for all TTV species was demonstrated in the range from 10 to 50*106 copies per reaction, with lower limit of detection as 1 copy per reaction.

Combining pre-transplant CYP3A5 genotyping with TTV monitoring post-transplant may further enhance immunosuppression management for kidney transplant patients.

Advancing translational cancer research: a unified DNA and methylation NGS workflow for cfDNA and gDNA from FFPE and FF tissues

POSTER 110

Grace Zhao[1], Heng Wang1, Bonita Lam1, Semen Leyn1, Jingga Inlora1, Xiaomu Chen1, Aswati Aravind1, Margherita Corioni1, Michael Ruvolo1, Douglas Roberts1, Kyeong-Soo Jeong1, Shengrong Lin1, Yun Bao1

1 Agilent Technologies, Santa Clara, CA

In translational cancer research, multiomic profiling across diverse sample types within a single workflow is often a challenge. A unified workflow enables consistent, comparable results, streamlined laboratory operations, and automation. Agilent Avida target enrichment technology, optimized for low input materials like cell-free DNA (cfDNA), offers simultaneous DNA variant and methylation analysis. We examine the feasibility of using genomic DNA (gDNA) from formalin-fixed paraffin-embedded (FFPE) and fresh frozen (FF) tissues as input for Avida Duo workflow.

We analyzed 10 tumor FFPE gDNA samples with varying sample quality and 78 FF gDNA samples (48 tumor, 30 normal). After fragmentation, DNA was used for library preparation followed by sequential capture (Avida Duo workflow): first targeting 500+ genes using comprehensive Avida DNA cancer panels (Discovery or Onco LB Plus), followed by targeting differentially methylated regions (DMRs) using the Avida Methyl 3400 DMR Cancer panel. We observed robust recovery of unique molecule indexes. Oncogenic variants, including one at <5% allele frequency, were detected and confirmed by an orthogonal targeted sequencing method. The methylation status for each DMR demonstrated cancer-related patterns.

For cfDNA, we examined 11 samples (5 normal, 6 cancer) using 1 mL of plasma. cfDNA yields and fractions were assessed using Agilent TapeStation. Libraries were prepared using the Avida Duo workflow and enriched with the Avida DNA Expanded Cancer and Avida Methyl 3400 DMR Cancer panels. Strong average coverage was achieved, enabling the detection of oncogenic variants with allele frequencies as low as ~0.2%. We observed excellent linearity between cfDNA input and molecule recovery, and clear separation of methylation indexes between cancer and normal groups.

We successfully demonstrated that plasma cfDNA, FFPE gDNA, and FF gDNA can all be analyzed using the same Agilent Avida chemistry and unified workflow for simultaneous DNA and methylation targeted analysis. The Avida Duo workflow (library preparation with target enrichment for DNA and methylation) can be completed within one day, without protocol modifications based on sample type, marking a significant advance for integrated comprehensive genomic and epigenomic profiling in translational cancer research.

For Research Use Only. Not for use in diagnostic procedures.

PR7001-4658

An enhancer deletion at APOE is protective against Alzheimer's disease but not lipid-related traits in African Americans

POSTER 111

Julianna N. Brutman[1], Tina Busald1, Evangelos Nizamis1, Eli J. Kaufman1, Eugene Lin2,3, Suman Jayadev1,4, Ellen M. Wijsman1,3, Elizabeth E. Blue1,5,6, Paul N. Valdmanis1,2,6

1 University of Washington School of Medicine, Division of Medical Genetics, Seattle, WA
2 University of Washington, Department of Genome Sciences, Seattle, WA
3 University of Washington, Department of Biostatistics, Seattle, WA
4 University of Washington School of Medicine, Department of Neurology, Seattle, WA
5 University of Washington, Institute for Public Health Genetics, Seattle, WA
6 Brotman Baty Institute, Seattle, WA

Common variants in APOE are strong risk factors for Alzheimer's disease (AD) and lipid-related traits like hyperlipidemia and hypercholesterolemia and these groups of phenotypes have historically been considered intrinsically linked. Risk of developing AD though varies greatly by genetic ancestry with greater risk in individuals of European rather than African ancestry for APOE-E4/E4 genotype carriers. To uncover potential protective alleles in cis with APOE-E4, we analyzed phased long-read APOE haplotypes and identified a 19bp deletion present in 60% of African American APOE-E4 homozygotes. The deletion is ~1.1kb distal to the APOE 3'UTR in a SPI1/PU.1 transcription factor binding site. SPI1 variants are also risk factors for AD and SPI1 drives development of microglia – a key cell type implicated in AD. Performing a phenome-wide association study (PheWAS) analysis in the NIH All of Us dataset revealed that after adjusting for APOE-E4 and APOE-E2, the odds ratio of developing AD for 19bp deletion carriers versus controls was 0.56 (p=0.0058). This protection was comparable to that afforded by the APOE-E2 variant (OR=0.55; p=4.88E10-6). Remarkably, deletion carriers still had significantly elevated relative risk for hyperlipidemia (OR=1.18, p=3.55E-13) and hypercholesterolemia (OR=1.24, p=2.88E-10) at levels similar to rs429358, the variant defining APOE-E4. PheWAS analysis for other variants between APOE and APOC1 exhibited risk exclusively for neurological or lipid-related traits thereby genetically uncoupling these groups of phenotypes at APOE. The deletion reduced AD odds ratio 8-fold compared to ancestry-matched individuals from the Alzheimer's Disease Sequencing Project without the deletion, confirming its role as a protective variant for AD. Analyzing individuals with local African ancestry at APOE revealed that the 19bp deletion was significantly protective against AD (OR=0.70; 95% CI 0.59-0.85; p=0.0002). The deletion also delayed AD onset by 3 years in APOE-E4/E4 cases with local African ancestry at APOE (p=0.01 by log-rank Mantel-Cox test). Functional assays revealed that the 19bp deletion abolished SPI1 repression at APOE and impacted APOC1 expression. Collectively, these findings uncover a long sought after protective allele at APOE in African Americans that mediates APOC1 expression, reducing relative AD risk.

Applying new tools and resources to analyze tandem repeat variation in short- and long-read rare disease cohorts

POSTER 112

Ben Weisburd[1,2], Egor Dolzhenko3, Hope Tanudisastro4-7, Anne O'Donnell-Luria1,2,8, Heidi L. Rehm1,2

1 Broad Institute of MIT and Harvard, Program in Medical and Population Genetics, Broad Center for Mendelian Genomics, Cambridge, MA
2 Massachusetts General Hospital, Harvard Medical School, Center for Genomic Medicine, Boston, MA
3 Pacific Biosciences of California, Menlo Park, CA
4 Garvan Institute of Medical Research, Centre for Population Genomics, Sydney, NSW, Australia
5 Murdoch Children's Research Institute, Centre for Population Genomics, Melbourne, VIC, Australia
6 University of New South Wales, Faculty of Medicine and Health, Sydney, NSW, Australia
7 University of Sydney, Faculty of Medicine and Health, Sydney, NSW, Australia
8 Boston Children's Hospital, Harvard Medical School, Division of Genetics and Genomics, Boston, MA

Tandem repeat (TR) expansions are known to cause over 60 rare diseases. Yet, despite significant progress over the past decade, our ability to discover novel causal TRs has been limited by the high cost and relatively low accuracy of genome-wide TR genotyping in short-read sequencing datasets. Additionally, TR discovery in both short-read and long-read datasets has suffered from the unavailability of accurate population allele frequency information and other annotations that are critical for interpreting and prioritizing TR variants.

To address these challenges, we first developed a comprehensive genome-wide TR catalog with rich annotations that include variation clusters, constraint scores, and TR allele frequencies from diverse populations. These allele frequencies cover over 4 million TR loci and are derived from thousands of long-read samples sequenced by the HPRC, 1kGP and AoU consortia. We also created a gnomAD-like browser that enables anyone to search, visualize, and download these annotations through an intuitive online interface at trexplorer.broadinstitute.org. Finally, we introduced substantial optimizations to the ExpansionHunter algorithm, reducing the cost of genotyping large TR catalogs in short-read sequencing datasets.

We then applied these new tools and resources to the analysis of genome-wide TR variation in 20,424 short-read and 1,269 long-read genomes from the GREGoR consortium and the Rare Genomes Project (RGP). Our pipeline used population allele frequencies as well as ExpansionHunterDenovo to select polymorphic TR loci in our catalog, and then applied the optimized ExpansionHunter as well as TRGT to genotype hundreds of thousands of TRs, yielding novel candidate TR loci that included a coding CAG locus within the EP400 gene.

A SXL3-targeting antisense oligonucleotides for treatment of Bainbridge-Ropers syndrome

POSTER 201

Yuji Nakamura[1,2], Isaac Tang1,2, Inna Shomer3, Nofar Mor3, Joseph G. Gleeson1,2

1 Rady Children's Hospital, Rady Children's Institute for Genomic Medicine, San Diego, CA
2 University of California San Diego, Department of Neurosciences, La Jolla, CA
3 n-Lorem Foundation, San Diego, CA
4 lllumina, Inc., Cambridgeshire, Cambridge, United Kingdom
5 Illumina, Inc., San Diego, CA

Additional Sex Combs-Like 3 (ASXL3) has been identified as a component of the chromatin-modifying Polycomb repressive deubiquitinase (PR-DUB) complex, which promotes transcription by counteracting polycomb-mediated transcriptional repression. Bainbridge–Ropers Syndrome (BRPS, MIM: 615485) is an autosomal dominant neurodevelopmental disorder resulting from certain heterozygous de novo truncating variants in the 3' region of the ASXL3 open reading frame (ORF). The clustering of 95% of reported BRPS pathogenic variants in the last two exons, whereas truncating variants in the 5' part of the ORF are seen in healthy individuals, is thought to reflect a likely toxic gain-of-function mechanism. Similarly, ASXL1 and ASXL2 also show 3' heterozygous de novo truncating variants correlated with other neurodevelopmental disorders, suggesting shared mechanisms. Here, we report the preclinical efficacy of antisense oligonucleotide (ASO)-mediated knockdown of ASXL3 in induced neuron (iNeuron) models, demonstrating partial correction of the disease-associated differential gene expression signature.

Induced pluripotent stem cells (iPSCs) from five BRPS patients and three healthy individuals were differentiated into iNeurons and cultured for 21 days. ASXL3 expression levels increased over time in culture, in line with its restricted expression in the maturing brain. Non-allele-selective ASOs targeting ASXL3 were applied to iNeurons via a free uptake method. The ASXL3-targeting ASOs suppressed ASXL3 levels by >90% after 6 days of treatment. Bulk RNA-sequencing in untreated patient-derived iNeurons revealed global dysregulation of gene expression compared with controls. Dysregulated genes showed significant enrichment for autism-related genes. Importantly, a substantial fraction of genes dysregulated in BRPS patients were significantly rescued by ASXL3-targeting ASOs. In contrast, treatment of control-derived iNeurons with ASXL3-targeting ASOs showed very few genes changed in expression, likely reflecting compensation by other ASXL family members.

Our approach provides critical proof-of-concept experimental data on the potential utility of ASOs to treat BRPS, and potentially other ASXL family member disorders using the same approach. Loss of ASXL3 function is tolerated in culture, supporting the feasibility of a non-allele-selective ASO therapeutic strategy, thus allowing all BRPS patients to be treated with a common ASXL3-targeting ASO.

Benchmarking large language models for predictive modeling in biomedical research with a focus on reproductive health

POSTER 202

Reuben Sarwal[1], Victor Tarca2*, Claire Dubin1, Nikolas Kalavros3,4, Gaurav Bhatti5, Sanchita Bhattacharya1, Atul Butte1,14, Roberto Romero6,7,8, Gustavo Stolovitzky9,10, Tomiko T. Oskotsky1,11, Adi L. Tarca5,12,13**, Marina Sirota1,14**

1 University of California, San Francisco, Bakar Computational Health Sciences Institute, CA
2 Huron High School, Ann Arbor, MI
3 New York University Langone Health, New York, NY
4 New York University Grossman School of Medicine, Division of Precision Medicine, Department of Medicine, New York, NY
5 Wayne State University School of Medicine, Center for Molecular Medicine and Genetics, Detroit, MI
6 National Institutes of Health, United States Department of Health and Human Services, Pregnancy Research Branch, Division of Obstetrics and Maternal-Fetal Medicine, Division of Intramural Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD
7 University of Michigan, Department of Obstetrics and Gynecology, Ann Arbor, MI
8 Michigan State University, Department of Epidemiology and Biostatistics, East Lansing, MI
9 New York University Grossman School of Medicine, Department of Pathology, New York, NY
10 New York University Langone Health, Biomedical Data Science Hub, New York, NY
11 University of California San Francisco, Department of Medicine, Division of Clinical Informatics and Digital Transformation, San Francisco, CA
12 Wayne State University School of Medicine, Department of Obstetrics and Gynecology, Detroit, MI
13 Wayne State University College of Engineering, Department of Computer Science, Detroit, MI
14 University of California San Francisco, Department of Pediatrics, San Francisco, CA
* co-first author
** co-senior author

Generative AI, particularly large language models (LLMs), is increasingly being used in computational biology to support code generation for data analysis. In this study, we evaluated the ability of LLMs to generate reproducible, high-quality R and Python code for predictive modeling tasks in genomics, leveraging standardized datasets from the several recent DREAM (Dialogue for Reverse Engineering Assessments and Methods) Challenges focused on maternal outcomes.

(continue to page 39)

(continued from page 38)

We assessed LLM performance across three predictive tasks derived from DREAM datasets: gestational age regression from gene expression, classification of preterm birth and early preterm birth from microbiome data, and gestational age regression from DNA methylation profiles. LLMs were prompted with task descriptions, data locations, and target outcomes. LLM generated code was run to fit prediction models and generate graphics. LLMs were evaluated based on their ability to generate executable code, produce valid visualizations, and achieve strong test set performance.

Among the eight LLMs tested, o3-mini-high, 4o, DeepseekR1 and Gemini 2.0completed at least one task without error. Overall, R code generation was more successful (14/16 tasks) than Python (7/16), attributed to the utility of Bioconductor packages for querying GEO data. OpenAI's o3-mini-high outperformed others, completing 7/8 tasks with test set performance matching or exceeding top-performing teams from one of the original DREAM challenges. These findings underscore the potential of LLMs to enhance exploratory analysis and democratize access to predictive modeling in genomics by automating key components of analysis pipelines and highlight the potential to increase research output conducting analyses of existing standardized datasets from public repositories.

Bridge Capture™ & Nicking Loop™: NGS library preparation methods for precision genomics and universal NGS readout

POSTER 203

Simona Adamusová[1,2], Anttoni Korkiakoski1,2, Nea Laine1, Anna Musku1, Tuula Rantasalo1, Jorma Kim1, Juuso Blomster1, Jukka Laine1,3,4, Tatu Hirvonen1, Manu Tamminen1,2, Juha-Pekka Pursiheimo1

1 Genomill Health Inc., Turku, Finland
2 University of Turku, Department of Biology, Turku, Finland
3 University of Turku, Department of Oncology, Turku, Finland
4 Turku University Hospital, Turku, Finland

Next-generation sequencing (NGS) has become increasingly affordable, with improved read quality and throughput. However, library preparation remains a major bottleneck. Traditional protocols, especially those designed for Illumina®, include steps that are now obsolete with newer sequencing platforms. These outdated processes introduce unnecessary complexity and bias, reducing the accuracy needed for modern applications in precision genomics and diagnostics. To address this, we developed two novel library preparation methods: Bridge Capture™ for targeted diagnostics and Nicking Loop™ for broad NGS applications. Both approaches eliminate outdated steps and use streamlined, PCR-free workflows to minimize bias and simplify adoption.

Bridge Capture™ was benchmarked against industry-standard methods, AmpliSeq™ CHPv2 (Illumina®) and Archer™ LIQUIDPlex™, using 24 matched contrived colorectal biospecimens mimicking ctDNA. It showed strong correlation in variant allele frequency (VAF) detection with AmpliSeq (R^2 = 0.988) and Archer (R^2 = 0.995), while also detecting the lowest VAFs. Reproducibility was high across different labs (R^2 = 0.979) and workflows (manual vs. automated, R^2 = 0.983), demonstrating its ease of integration into new lab environments, including automation platforms. Bridge Capture™ is platform-agnostic, cost-effective, fast, and sensitive, making it a robust solution for diagnostic applications.

Nicking Loop™ enables conversion of both single- and double-stranded DNA into circular single-stranded DNA (CssDNA), allowing for controlled, unbiased amplification. The method produces multiple DNA forms: CssDNA for circular-DNA platforms, linear DNA for traditional systems like Illumina®, and concatemers for long-read sequencers like nanopore. The power of this method is highlighted by direct sequencing of circular Nicking Loop™ libraries on the PacBio Onso™ platform, which achieved a Phred quality score of Q48, far exceeding the commonly accepted Q30 threshold.

Together, these technologies demonstrate that by rethinking library preparation to align with the capabilities of modern NGS platforms, we can dramatically improve workflow simplicity, data quality, and sequencing performance. Bridge Capture™ and Nicking Loop™ represent a new generation of library preparation methods designed to unlock the full potential of today's and tomorrow's sequencing technologies.

Calibrating evidence to improve variant classification for genomic medicine

POSTER 204

Anne O'Donnell-Luria[1,2], Daniel Zeiberg3, Haneen Abderrazzaq3, Vikas Pejaver4,5, Sarah L Stenton1,2, Himanshu Sharma4, Timothy Bergquist4, Mugdha Singh1,2, Emily A.W. Nadeau6, Alicia B. Byrne1, Marc S Greenblatt6, Steven M Harrison1,7, Sean V. Tavtigian8, Leslie G. Biesecker9, Steven E. Brenner10, Predrag Radivojac3

1 Broad Institute of MIT and Harvard, Program in Medical and Population Genetics, Cambridge, MA
2 Boston Children's Hospital, Harvard Medical School, Division of Genetics and Genomics, Boston, MA
3 Northeastern University, Khoury College of Computer Sciences, Boston, MA
4 Icahn School of Medicine at Mount Sinai, Institute for Genomic Health, New York, NY
5 Icahn School of Medicine at Mount Sinai, Department of Genetics and Genomic Sciences, New York, NY
6 University of Vermont, Department of Medicine and University of Vermont Cancer Center, Larner College of Medicine, Burlington, VT
7 Ambry Genetics, Aliso Viejo, CA
8 University of Utah School of Medicine, Huntsman Cancer Institute, Department of Oncological Sciences, Salt Lake City, UT
9 National Institutes of Health, National Human Genome Research Institute, Center for Precision Health Research, Bethesda, MD
10 University of California, Department of Plant and Microbial Biology and Center for Computational Biology, Berkeley, CA

Accurate classification of genetic variants as pathogenic or benign is crucial to the delivery of genomic medicine. The American College of Medical Genetics and Genomics and Association for Molecular Pathology, in collaboration with ClinGen, have developed and refined standards for the classification of genomic variation by evaluating population, computational, functional, and case data. These measures have substantially improved the quality and consistency of genomic results. The weighting of evidence in these assessments has been largely guided by expert opinion, which likely results in under- and over-weighting some evidence. There are now sufficient data available to quantitatively assess and define the strength of certain evidence types, from population data (gnomAD), clinically classified variants (ClinVar), and multiplex assays (MaveDB). For example, we found that some computational predictors can provide up to Strong (+4 points) evidence for missense variant pathogenicity/benignity when variants reach thresholds identified by a local probability-based calibration algorithm, while initial guidelines limited computational evidence to Supporting (+1 point).

(continue to page 42)

(continued from page 41)

This recommendation was endorsed by the ClinGen Sequence Variant Interpretation Working Group for use in clinical practice. We have expanded this to in-frame indels, which can reach up to Moderate (+2 points), and also to regional missense constraint, which can independently reach Moderate (+2) evidence. The increasing availability of high-throughput functional data is aiding the resolution of variants of uncertain significance (VUS). Using a multi-sample skew-normal mixture of distributions of functional (ClinVar) and population (gnomAD) variants, we applied a constrained expectation-maximization algorithm to calculate variant-specific evidence strengths for use in the clinic.

These approaches have increased the number of clinically informative classifications while also providing a rigorous basis for the quality of evidence applied, which should improve the reliability of variant classification. A validation study demonstrated that these approaches affect very few variants per individual with rare disease, consistent with the expectation that an individual's Mendelian condition typically arises from only 1 or 2 variants. Calibration of evidence approaches are improving the delivery of genomic medicine by resolving VUS, thereby increasing diagnostic yield.

Comparative canine-human multi-omics approach for discovering therapeutic drug combinations in human osteosarcoma

POSTER 205

Gregory Tawa[1], Amy K. Leblanc2, David Gerhold1, Matthew Breen3, Craig Thomas1, Christina Mazcko2, Scott Hoyt1, John Braisted1, Gavrila Wicaksono1, Shan Huang2, Ling Ren2, Crystal McKnight1, Kelli Willson1, Carleen Klumpp-Thomas1, David Holland1, Xiaohu Zhang1

1 National Institutes of Health, Therapeutic Development Branch, Division of Preclinical Innovation, National Center for Advancing Translational Sciences, Rockville, MD
2 National Institutes of Health, National Cancer Institute, Center for Cancer Research, Comparative Oncology Program, Bethesda, MD
3 North Carolina State University, Department of Molecular Biomedical Sciences, College of Veterinary Medicine, Raleigh, NC

Osteosarcoma is diagnosed in fewer than 400 children annually in the United States, limiting opportunities for therapeutic development for this rare childhood cancer. In contrast, osteosarcoma is common in canines, affecting over 10,000 dogs per year in the U.S. Growing evidence indicates that canine osteosarcoma shares critical clinical and genomic features with the human disease. The high incidence of osteosarcoma in dogs, combined with these cross-species similarities, creates a significantly expanded opportunity for human therapeutic discovery.

To capitalize on this, we developed an integrated experimental and bioinformatics pipeline that generates and analyzes mRNA and genome sequencing data across multiple species. This pipeline identifies disease-associated genes shared among species and prioritizes potential drug targets and therapeutic combinations based on these shared genes. Candidate genes are validated through proteomics to confirm that they are expressed at the protein level in tissues. Approved drugs that modulate these disease-associated proteins are identified, and drug combinations are designed such that each drug targets a distinct protein, enhancing the potential for synergy.

These drug combinations are evaluated in canine cancer cell lines, and those demonstrating strong efficacy and synergy are selected for further testing in mouse xenograft models. The ultimate goal is to translate these findings into in vivo studies in both canines and humans. This poster presents key results from applying this pipeline over the past year. Our approach highlights the relevance of canine models to human cancer research and offers new opportunities to investigate cancer biology and therapeutic strategies across species.

Development and validation of microarray-based methylation classification for central nervous system tumor for clinical testing

POSTER 206

Marco L. Leung[1-3], Alexandra Evans1, Kareesma Parbhoo1, Diana Thomas2,4, Daniel Boue2,4, Christopher Pierson2,4, Ben Kelly1, Catherine Cottrell1-3, Richard Wilson1,3, Elaine Mardis1,3

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

Methylation profiling has demonstrated both diagnostic and clinical utilities in aiding diagnoses in cancers, especially in central nervous system (CNS) tumors, however its adaptation in the clinical laboratory setting is difficult due to the unavailability of existing classifiers for inhouse usage. Moreover, the bioinformatic development of a methylation classifier is challenged by the need for comprehensive acquisition of tumor subtypes, and the variable methylation datasets generated by different microarray platforms. Here, we present a newly developed classifier software that classifies CNS tumors based on methylation profiles developed at Nationwide Children's Hospital Institute for Genomic Medicine (IGM). The IGM classifier was developed using 3979 CNS samples with 109 CNS subtypes based on 10,000 probes that are present in three Illumina methylation microarray platforms (Infinium Human Methylation 450K BeadChip, Infinium MethylationEPIC v1.0 850k BeadChip, and Infinium MethylationEPIC v2.0 930k BeadChip). We demonstrated that our software could provide CNS subtypes with high confidence scores in 83% of 230 validation samples, and high concordance with the German Cancer Research Center CNS v12.8 Classifier. To further understand the clinical utilities of this methodology, methylation profiling results in refinement of diagnosis in 33.3% of patients, and change in diagnosis in 5% of patients. In summary, we expect this classifier to be an invaluable tool in aiding pathologists and oncologists to better refine diagnoses, thus improving patient outcome.

Development of a CRISPRa therapy to rescue CTNND2 haploinsufficiency

POSTER 207

Dianne Laboy Cintrón[1,2], Nadav Ahituv1,2

1 University of California San Francisco, Department of Bioengineering and Therapeutic Sciences, San Francisco, CA, USA
2 University of California San Francisco, Institute for Human Genetics, San Francisco, CA, USA

Cri du chat syndrome (CdCs) is a rare neurodevelopmental disorder caused by a deletion in the short arm of chromosome 5. CdCs is characterized by a high-pitched cry, poor growth, developmental delay, and dysmorphic features. Within the deleted region, CTNND2 is a haploinsufficient gene that has been demonstrated to contribute to the neurological phenotype observed in CdCS. Previous studies in our laboratory have shown that CRISPR activation (CRISPRa) can be used to rescue haploinsufficiency. We used this approach to rescue CTNND2 expression as a potential therapeutic strategy for CdCs. We designed and screened guide RNAs (gRNAs) for their ability to upregulate CTNND2. We identified a lead candidate gRNA that robustly increases CTNND2 gene expression. To further evaluate the lead candidate gRNA, we established patient-derived neurons and implemented a dual AAV9 delivery of the CRISPRa system to assess CTNND2 upregulation. This project lays the groundwork for therapeutic development targeting CTNND2 in CdCS.

Development of an anti-sense oligonucleotide therapeutic targeting RhoBTB2-related epileptic encephalopathy

POSTER 208

Thomas J. LaRocca[1], Sara Emami1, Devon Bonner1, Raquel Alvarez1, Richard A. Coca1, Jonathan A. Bernstein1, Kala Nair1, Juliet Knowles1, KC Kent Lloyd1, Matthew T. Wheeler1

1 Stanford University School of Medicine, Department of Medicine, Palo Alto, CA

Pathogenic variants within RHOBTB2 describe the autosomal dominant developmental and epileptic encephalopathy 64 [DEE64] with several mutations reported to cause overaccumulation of RHOBTB2 due to impaired protein degradation. We hypothesized that the development of an anti-sense oligonucleotide (ASO) with the goal of reversing accumulation of RhoBTB2 may benefit patients with DEE64.

Following IRB approved protocol and consent, research genome testing of a 23-month-old male with global developmental delay, epilepsy, dystonic posturing, and negative exome testing revealed a heterozygous de novo pathogenic variant in RHOBTB2 c.1448G>A, p.Arg483His. Patient-derived skin fibroblasts were reprogrammed to generate induced pluripotent stem cell (iPSC) lines, utilizing CRISPR with HDR to create isogenic control iPSCs. We developed a patient registry to facilitate natural history data collection. Patient derived cortical organoids indicated impaired proliferation and increased evidence of apoptosis compared to CRISPR corrected isogenic control ogranoids. A RhoBTB2 p.R461H mouse locus in C57BL/6 strain was developed to emulate the human RhoBTB2 p.R483H variant using CRISPR with homology directed repair (HDR). Mice were phenotyped using neurodevelopmental assessments, in vivo EEG, and brain MRI/MRS. The humanized mouse model of RhoBTB2 p.R483H-related disease recapitulates key phenotypes of the human disorder including impaired growth, stereotyped movements, impaired REM sleep and increased duration of pilocarpine induced seizures in heterozygous mice compared to wild type mice (300sec +/- 100sec vs 75sec +/- 15). MRI and spectroscopy revealed a reduction in GABA signal (61.4mM vs 64.1mM, n=5) with no significant differences in fractional anisotropy (0.34 vs 0.35), or mean diffusivity (0.66 vs 0.64) compared to wild-type mice in Mescher-Garwood Point Resolved Spectroscopy (MEGA-PRESS) and diffusion tensor imaging and respectively. Finally, three ASO sequences targeting RhoBTB2 c.1448G>A led to an 80% reduction in RhoBTB2 compared to scrambled control ASOs.

These models share mechanistic and key neurodevelopmental phenotypes of DEE64. Through parallel model systems of the RhoBTB2 R483H pathogenic variant, we designed and tested therapeutic candidates for DEE64. Further work is on-going to understand RhoBTB2 neurobiology and ASO optimization, safety and efficacy.

Direct RNA transcriptomic and epitranscriptomic profiling of human brain cortex using the Dogme computational pipeline

POSTER 209

Elnaz Abdollahzadeh[1,2], Ali Mortazavi1,2

1 University of California, Irvine, Department of Developmental and Cell Biology, Irvine, CA
2 University of California, Irvine, Center for Complex Biological Systems, Irvine, CA

The role of RNA modifications in health and disease has been difficult to study because of the difficulty in measuring modification levels using traditional sequencing techniques. Oxford Nanopore Technologies (ONT) enables direct detection of RNA and DNA modifications from unamplified nucleic acids, providing unique insights into epigenetic and epitranscriptomic regulation. However, rapid updates to ONT basecalling models and evolving modification-detection tools pose challenges for reproducible analysis. To address these, we developed Dogme, a reproducible, Nextflow-based pipeline that automates ONT data processing, including basecalling (Dorado), alignment (minimap2), transcript annotation, modification detection (modkit), and transcript quantification (LR-Kallisto), across direct RNA (dRNA), complementary DNA (cDNA), and genomic DNA (gDNA) datasets. Dogme supports diverse RNA modifications detectable by Dorado, including N6-methyladenosine (m6A), 5-methylcytosine (m5C), pseudouridine (Psi), inosine (I), and 2'-O-methylation (Nm), as well as DNA methylation.

We applied Dogme to ONT PromethION dRNA sequencing of postmortem human cortex tissue from two female Alzheimer's disease (AD) patients and two age-matched controls from the ENCODE AD collection. We identified significant differences in multiple RNA modifications, with AD samples showing increased m6A and pseudouridine, along with distinct patterns in m5C, inosine, and Nm. We identified modification sites present in all four samples as well as a subset only found in AD samples, and we are currently analyzing the potential impact of these differences.

Distributed artificial intelligence for genomic and clinical data from multiple institutions

POSTER 210

Tsung-Ting Kuo[1,2]

1 Yale School of Medicine, Department of Biomedical Informatics and Data Science, New Haven, CT
2 Yale School of Medicine, Department of Surgery, New Haven, CT

To improve the performance of genomic/medical Artificial Intelligence (AI) approaches and to identify gene/medication-outcome associations for diseases, it is critical to integrating data "vertically" (e.g., a patient has genomic data in an institution and clinical data in another one) or "horizontally" (e.g., when multiple patients have the same type of genomic and/or clinical data from different institutions), from multiple institutions. Compared to "siloed" AI model learning, vertical data integration can include more covariate information, while horizontal one can increase sample size. Intuitively, the data can be transferred to a centralized repository for model learning, yet such a dissemination of genomic/medical data may present privacy risks such as re-identification. Existing Privacy-Enhancing Technologies (PETs) such as federated learning allow institutions to disseminate only aggregated AI models without exchanging patient-level data. However, the federated model learning process in general requires a moderating central server, which could present an undesirable single-point-of-control. To overcome this issue, we developed a completely distributed AI approach across institutions with the following desirable technical features: (a) integrating both vertically- and horizontally-partitioned data without patient-level record dissemination; (b) providing a computationally-fair and fully-distributed cross-institutional AI learning; (c) leveraging community-supported distributed platforms and centralization-equivalent methods with an immutable ledger; and (d) building scalable, transparent, and trustworthy distributed AI models. The use of genomic and clinical data from multiple institutions can be supported by our models. Our framework can facilitate collaborative privacy-preserving modeling across multiple genomic and healthcare institutions with vertically- or horizontally-partitioned data as a completely distributed AI algorithm.

Acknowledgements. The authors were funded by the U.S. National Institutes of Health (NIH) (R01EB031030). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

DNA language models support the analysis of ancestry-specific APOE regulation and Alzheimer's disease risk

POSTER 211

Rui Zhu[1], Xiaopu Zhou2, Stephen W. Scherer2,3, Lucila Ohno-Machado1

1 Yale University, Department of Biomedical Informatics & Data Science, School of Medicine, New Haven, CT
2 The Hospital for Sick Children, Program in Genetics and Genome Biology, Toronto, Ontario, Canada
3 University of Toronto, McLaughlin Centre and Department of Molecular Genetics, Toronto, Ontario, Canada

Large language models (LLMs) trained on genomic sequences encode rich information about evolutionary constraints, yet their utility for interpreting complex haplotype architectures remains largely unexplored. Meanwhile, conventional haplotype analyses often rely on large cohorts to capture subtle effects, limiting their scalability. Here, we leverage Evo2, a state-of-the-art DNA-LLM, to perform ancestry-aware haplotype analysis in Alzheimer's Disease (AD). We applied Evo2 to 240 fully phased human genome assemblies from the Human Pangenome Reference Consortium, each annotated with high-resolution local ancestry and APOE E2/E3/E4 genotypes, to compute individual haplotype scores.

We studied the association of Evo2 haplotype scores and regulatory impact: across ancestry groups, mean haplotype scores were tightly correlated with mean APOE expression ($r = 0.86$, $p < 2\times10^{-16}$). Within individuals ($n = 33$), haplotypes with a low Evo2 score predicted low APOE transcript levels ($p < 0.05$), indicating that Evo2 captured cis-regulatory potency despite no exposure to transcriptomic data.

Additionally, Evo2 haplotype scores recapitulated known ancestry-specific disease risk patterns. Haplotypes carrying the APOE E4 allele were enriched in the high-score group, whereas E3 haplotypes clustered centrally and E2 haplotypes were enriched in the low-score group. Among E4 tagging haplotypes, median haplotype scores decreased monotonically across ancestries, from East Asian (1.33e-4), which was the highest, to African (-2.52e-4), which was the lowest, paralleling reported Alzheimer's disease odds ratios; notably, Evo2's APOE-E4 tagging haplotype scores were highly correlated with these odds ratios ($r = 0.92$).

Our findings demonstrate that LLM-based haplotype scoring may reveal regulatory and ancestry-linked risk differences without the large sample sizes required by classical haplotype analyses. This approach may therefore offer a scalable tool for precision health applications such as variant prioritization and disease stratification, in an ancestry-aware manner.

*Partly funded by RM1HG011558

Electronic genome mapping bioinformatics pipeline for repeat expansion detection

POSTER 212

Shuk Shukor[1], Dong Zhang2, Xu Tan2, Reger Mikaeel2, Lindsay Schneider2, Kaylee Mathews2, Jason Hughes2, Michael Kaiser2, John Thompson2

1 Hitachi High-Tech America, San Diego, CA
2 Nabsys, Providence, RI

Background: Repeat expansion disorders are caused by abnormally long short tandem repeats, with pathogenicity often linked to expansion size and age of onset. For example, CGG expansions in FMR1 lead to Fragile X syndrome. Traditional diagnostics like Southern blotting and PCR are labor-intensive, low-throughput, and limited in scalability. While next-generation sequencing (NGS) offers genome-wide potential, it struggles with GC-rich, repetitive regions—especially large expansions.

Methods: We applied electronic genome mapping (EGM), a solid-state nanochannel technology, for repeat expansion detection. DNA from cell lines was prepared via a dual-enzyme protocol and injected onto the OhmX™ Analyzer for data collection. Repeat expansions were analyzed using SV-Verify™ (SVV) repeat expansion pipeline.

Results: To evaluate the capability of detecting repeat expansions, we analyzed the FMR1 repeats relevant to Fragile X syndrome using six Coriell cell lines with known repeat sizes and six control samples. We observed the expected expansion alleles in the Coriell cell lines, with sizes consistent with annotation and with the largest expansion being almost 600 copies. The control samples had repeats below the full expansion cutoff.

Conclusion: EGM provides a scalable and precise solution for genome-wide detection of repeat expansions. This pipeline currently supports detection in FMR1 and is being extended to other pathogenic loci, including DMPK, FXN, and C9orf72, enabling the development of a multiplexed diagnostic panel for repeat expansion disorders.

Engineering “living” materials with Bacillus subtilis as a biocontrol agent to combat antimicrobial resistance and disease transmission in built environments

POSTER 301

Kathleen L. Furtado[1,2], Lin Huang3, William Brakewood4, Maxwell Neal5, Jacob Hizon2, Shivam Singhal2, Mariana Salas Garcia1,2, Joshua Tran1,2, Karsten Zengler1,5,6, Michael Betenbaugh4, Shaochen Chen3,5, Jack A. Gilbert1,2,6

1 University of California San Diego, Department of Pediatrics, La Jolla, CA, USA
2 University of California San Diego, Scripps Institution of Oceanography, La Jolla, CA, USA
3 University of California San Diego, Department of Chemical and Nano Engineering, La Jolla, CA, USA
4 Johns Hopkins University, Department of Chemical and Biomolecular Engineering, Baltimore, MD, USA
5 University of California San Diego, Department of Bioengineering, La Jolla, CA, USA
6 University of California San Diego, Center for Microbiome Innovation, La Jolla, CA, USA

Antimicrobial resistance (AMR) is an increasingly deadly global threat. Of AMR pathogens, methicillin-resistant Staphylococcus aureus (MRSA) remains a serious concern, as progress to reduce new infections has slowed. AMR pathogens are often acquired via contact with contaminated surfaces, and as standard cleaners promote resistance, there is an urgent need for new tools to reduce pathogen burden. Bacillus subtilis is a promising biocontrol agent for built environments as it is generally regarded as safe, often applied as a probiotic, genetically tractable, and its spores can persist in harsh environments. It also produces numerous antimicrobial compounds with varied mechanisms of action against pathogens, including MRSA. Our goal is to develop biologically active surfaces containing B. subtilis spores as alternative approach to combat AMR. We hypothesize that embedding surface materials with B. subtilis spores will provide a stable form of biocontrol: upon germination, B. subtilis is poised to effectively outcompete and inhibit common surface-associated pathogens. We assessed the competitive fitness of multiple wildtype B. subtilis isolates in co-cultures with MRSA and our results suggest that MRSA inhibition varies by strain. Among all strains tested, B. subtilis TH035, a soil isolate, had the highest competitive indices against MRSA. These results corresponded to predictions from a recently established pan-genome-scale metabolic model for B. subtilis, which determined that the inhibitory potential of strain TH035 ranks within the top 10% of the 491 strains modeled. We then tested the antimicrobial activity of this strain in 3D-printed surface material prototypes, fabricated by digital light processing bioprinting. We confirmed that B. subtilis spores germinate within these materials, yielding >99% germination efficiency under nutrient-rich conditions. We also observed a significant, 7.46-fold reduction in MRSA growth in media surrounding spore-embedded materials relative to control materials without spores. In future work, we will explore the transcriptomic and metabolic responses of B. subtilis to MRSA in competitions and use genetic engineering to further optimize B. subtilis competitive fitness. Ultimately, we hope to develop a precision tool to reduce pathogen burden and manipulate built environment microbiomes in ways which promote human health.

Evaluating the Impact of Clinical Phenotyping on Causal Gene Prioritization in Rare Diseases

POSTER 302

Jordan L. Cahoon[1,2], Stephen B. Montgomery1,2,3 , Emily Alsentzer1

1 Stanford University, Department of Biomedical Data Science, Palo Alto, California, United States
2 Stanford University, Department of Pathology, Palo Alto, California, United States
3 Stanford University, Department of Genetics, Palo Alto, California, United States

Over 300 million patients worldwide live with a rare disease, with each disease having distinct clinical presentations and genetic causes. The low prevalence and the heterogeneous clinical presentations of each disease make diagnosis challenging. For patients without a definitive diagnosis, geneticists integrate sequencing data and detailed clinical phenotypes to identify causal disease mechanisms. This requires aligning a patient's genomic and phenotypic features with curated databases and previously diagnosed cases, which relies on capturing the clinical presentation with accuracy and sufficient granularity. However, phenotypes may evolve over the course of disease and documentation may vary among clinicians; incomplete and inconsistent phenotypic data can impede accurate comparisons and prolong the diagnostic process.

To examine the impact of phenotyping variability on diagnosis, we quantified variation in clinical phenotyping across two rare disease cohorts: Genomics Research to Elucidate the Genetics of Rare Diseases (GREGoR) and the Undiagnosed Diseases Network (UDN). The number of recorded phenotypes varied (GREGoR: $\mu = 7.8$, $E = 7.0$; UDN: $\mu = 18.6$, $E = 14.6$), as did their specificity, measured by Human Phenotype Ontology depth (GREGoR: $\mu = 6.4$, $E = 1.0$; UDN: $\mu = 6.6$, $E = 0.7$). When matching for disease, UDN participants had on average a 450% increase in number of phenotypes and 13% increase in specificity of phenotypes than those of GREGoR participants. We then assessed how this variation impacts causal gene ranking. Our results indicate that the quality of clinical phenotypes potentially correlates with better gene prioritization, highlighting the need for standardized, comprehensive phenotyping to accelerate rare disease diagnosis. This work will inform next steps for enhancing phenotyping practices in the GREGoR consortium, ensuring the long-term utility of this national resource for discovery and diagnosis.

Evaluation of Bridge Capture™ technology for mutation profiling in liquid biopsies of metastatic colorectal cancer patients

POSTER 303

Anttoni Korkiakoski[1,2], Simona Adamusová1,2, Aparna Ganesan3,4, Anna Musku1, Tuula Rantasalo1, Nea Laine1, Emma Andersson5, Emerik Österlund6,7, Ali Ovissi8, Päivi Halonen9, Tatu Hirvonen1, Jorma Kim1, Jukka Laine1,10, Antti Silvoniemi10,11, Heikki Minn10, Juuso Blomster1, Anna-Kaisa Anttonen5, Soili Kytölä5, Pia Osterlund9,12,13, Juha-Pekka Pursiheimo1, Pirjo Nummela3,4, Manu Tamminen1,2, Ari Ristimäki3,4

1 Genomill Health Inc, Turku, Finland
2 University of Turku, Department of Biology, Turku, Finland
3 University of Helsinki, Applied Tumor Genomics Research Program, Helsinki, Finland
4 Helsinki University Hospital and University of Helsinki, Department of Pathology, Helsinki, Finland
5 Helsinki University Hospital and University of Helsinki, Department of Genetics, Helsinki, Finland
6 Helsinki University Hospital and University of Helsinki, Department of Surgery, Helsinki, Finland
7 Uppsala University, Department of Immunology, Genetics and Pathology, Uppsala, Sweden
8 Helsinki University Hospital and University of Helsinki, Department of Radiology, Helsinki, Finland
9 Helsinki University Hospital and University of Helsinki, Department of Oncology, Helsinki, Finland
10 Turku University Hospital, Department of Oncology, Turku, Finland
11 Turku University Hospital and University of Turku, Department of Otorhinolaryngology, Turku, Finland
12 Tampere University Hospital and University of Tampere, Department of Oncology, Tampere, Finland
13 Karolinska University Hospital and Karolinska Institute, Department of Oncology/Pathology, Stockholm, Sweden

Colorectal cancer (CRC) is the second leading cause of cancer-related deaths, often presenting at an advanced stage with significant molecular heterogeneity. This is the first study to evaluate the performance of a novel next-generation sequencing (NGS)-based Bridge Capture technology for mutation profiling and minimal residual disease detection in circulating tumor (ct)DNA from metastatic colorectal cancer (mCRC) patients. Its performance was compared to those of droplet digital PCR (ddPCR), Ion AmpliSeq Cancer Hotspot Panel v2, and Idylla ctKRAS Mutation Assay. Eighty serial plasma samples from ten mCRC patients were analyzed by Bridge Capture and ddPCR, demonstrating a very strong correlation in variant allele frequency (VAF) values (rs = 0.86). The concordance of Bridge Capture with ddPCR (kappa = 0.70) and Idylla (kappa = 0.79) showed substantial agreement. A subset of samples (n = 10) was analyzed using the Ion AmpliSeq NGS-panel and both methods identified 15 driver mutations with strong correlation of VAF values (rs = 0.74). Additionally, Bridge Capture identified several oncogenic mutations beyond those detected by Ion AmpliSeq, highlighting its comprehensive profiling capability. The scalability of Bridge Capture was validated using an expanded panel and synthetic DNA targets, showing a strong linear correlation between observed and expected VAF values. This study demonstrates the scalability and accuracy of the Bridge Capture platform, and its potential to enhance mutation detection and clinical decision-making using ctDNA samples from patients with mCRC.

Exons encoding protein domains are associated with increased splice site recognition

POSTER 304

Laurens van de Wiel[1-3], Federico Ferraro1-4, GREGoR Consortium, Stephen B. Montgomery2,3, Matthew T. Wheeler1

1 Stanford University, School of Medicine, Division of Cardiovascular Medicine, Stanford, CA
2 Stanford University, School of Medicine, Department of Genetics, Stanford, CA
3 Stanford University, School of Medicine, Department of Pathology, Stanford, CA
4 University Medical Center Rotterdam, Erasmus MC, Department of Clinical Genetics, Rotterdam, The Netherlands

The human genome encodes 19,209 protein-coding genes and through alternative splicing gives rise to at least 47,778 protein isoforms. Protein domains constitute ~46% of human protein-coding genes. Unexpectedly, we previously found that rare genetic variation in protein domains is associated with outliers in splicing, methylation, gene-, and protein expression, a 4-9 fold stronger than other protein-coding positions. We here further evaluate the relationship between protein domains and splicing.

We created one-to-one mappings between reference genome annotation GENCODE v45 and UniProtKB/Swiss-Prot protein sequences. This led to 45,778 transcripts of 19,209 protein-coding genes matched to 33,633 unique protein sequences. These transcripts consist of 501,831 exons, of which 457,692 are protein-coding. Using PfamScan, we annotated this sequence set with 7,984 Pfam v37.2 domain families and 108,521 domain occurrences. These domain regions are encoded over 288,849 (63.1%) of the protein-coding exons. We extracted 466,932 genomically unique splice sites across all exons. Next, we used SpliceAI scores to analyse these splice sites in four groups; 51,276 (11.0%) in non-coding exons, 155,793 (33.4%) in exons that do not encode protein domains, 138,382 (29.6%) in exons that partially encode a protein domain, and 121,481 (26.0%) in exons that fully encode protein domain regions. We observed the median and median absolute deviation (MAD) of SpliceAI scores were 0.82 [MAD=0.37] for splice sites in non-coding exons, 0.98 [MAD=0.15] for exons without protein domains, 0.99 [MAD=0.10] for exons partially encoding protein domains, and 0.99 [MAD=0.08] for exons fully encoding protein domains.

Our findings indicate that protein-coding exon boundaries are, as expected, predicted with higher confidence to be splice sites. However, the lower MAD for the domain-encoding groups highlights that this confidence is more consistent for exons that encode protein domains, suggesting that protein domain architecture may inform alternative exon usage and could be used to characterize novel transcripts.

Federated data querying implementation for improving performance and clinical validation of genomic newborn sequencing

POSTER 305

Meredith Wright[1], Jeremy Leipzig2, Yupu Liang3, William Mowrey3, Valentina Nocco4, Chad Krilow2, Danny Oh1, Hugh Kwon1, Kevin Hall4, Rebecca Reimers1, Thomas DeFay3, Stephen Kingsmore1

1 Rady Children's Institute for Genomic Medicine, San Diego, CA
2 TileDB
3 Alexion AZ
4 Illumina

Genome sequencing has the potential to expand newborn screening (gNBS) to hundreds of rare childhood diseases that have effective treatments. Individual clinical trials of gNBS will be under-powered to evaluate the analytic performance, clinical utility, or cost effectiveness of gNBS. Some pathogenic (P) and likely pathogenic (LP) variants in gNBS genes have low penetrance or expressivity in population screening, leading to low positive predictive value of gNBS. Different causal genotypes and frequencies in different genetic ancestries is also a concern. A potential solution is data sharing between gNBS clinical trials the overcomes logistical, regulatory, and privacy concerns around data sharing. Query federation, where a query is shared rather than data, returns aggregated, deidentified results, that can overcome these concerns. To demonstrate the feasibility of query federation, we analyzed the 3202 1KGP WGS VCFs in three different database stores (TileDB, DNAnexus, Illumina Connected Analytics), in different geographical regions, and three different organizations. The BeginNGS query was designed to search individual VCFs for >40,000 variants associated with 412 disorder-gene pairs and return aggregated diplotype counts. Following optimization, identical results were obtained allowing for additional assessment of false positive BeginNGS variants. After query iteration, the aggregated positive diplotype count with the revised block list was 0.038 positives per subject. Multiple security tests were performed to ensure only data for the agreed upon fields could be queried and returned. Aggregated output data from the three stores were compared to test the accuracy of the returned data and modify the query scripts to be more platform-agnostic. This study demonstrated the ability to generalize the BeginNGS query to different data sets hosted by different cloud platforms and organizations. The BeginNGS query and test data are publicly available on our github site (https://github.com/rady-childrens-genomics/beginngs-partner-onboarding). Next we will enhance integration and automation across systems, expand the federated model to be used on a wider set of public reference genomes and collaborators' private genomes to improve the specificity and sensitivity of our gNBS test. Our goal is to expand federated queries to all P, LP and VUS in ~5,000 genes associated with single locus genetic disorders and to multiple genetic ancestries.

Genomic sequencing for undiagnosed neurological disorders: the Rady Children's Precision Medicine for Rare Disease Program

POSTER 306

Kristen Wigby[1,2], Jerica Lenberg1, Matthew Bainbridge1, Corrine Blucher1, Hung Nguyen1, Hannah Oppenheim3, Jennifer Schleit1, Stephen Kingsmore1, Annette Feigenbaum1,2, Jennifer Friedman1-3

1 Rady Children's Institute for Genomic Medicine, San Diego, CA
2 University of California, San Diego, Department of Pediatrics, La Jolla, CA
3 University of California, San Diego, Department of Neurosciences, La Jolla, CA

Genomic sequencing (GS) provides unparalleled opportunities for diagnosis and treatment of rare genetic disorders. Though GS has increased in genetics and neurology clinics several challenges remain including: 1) insurance barriers to GS, 2) diagnostic approach following non-diagnostic GS in patients with compelling phenotypes, 3) clinical correlation of multiple variants in complex patients, 4) delineation of candidate genes to a patient's phenotype, 5) lack of therapies for rare genetic diseases.

The Precision Medicine for Rare Disease Program (PMRDP) is a multi-dimensional collaboration of physicians, genetic counselors and scientists assembled with the goal of providing diagnoses and potential therapeutic options to individuals with rare and difficult to diagnose genetic disorders. The PMRDP consists of the patient facing Precision Medicine Clinic (PMC) that functions using a multi-disciplinary clinical care model, and the behind-the-scenes collaboration of physicians, scientists, counselors in addition to collaborating research labs. Where applicable, short read GS is utilized to circumvent the diagnostic odyssey. When non-diagnostic, on a research basis, additional in-depth genomic review and non-clinical omics tools including long-read sequencing, methylation studies, RNA sequencing, metabolomics are employed. Variants in candidate or established genes potentially relevant to the patient's phenotype are evaluated with functional studies – often through collaborative research efforts.

PMC has evaluated 295 patients (age 0.7 to 34 years), many of whom had extensive prior testing. As of June 2025, of 275 children without a prior diagnosis, a genetic diagnosis was achieved for 80 medically complex children (29%) and a partial or possible diagnosis was identified in 85 other children (31%). Diagnoses were achieved by multiple methods including short read GS, other clinical testing, bioinformatic re-analysis of existing GS data, other 'omics assays, and research collaborations. Research collaborations were initiated for 43 patients to investigate novel genes and/or variants and to further delineate phenotypes of ultra-rare disorders. Targeted treatment trials have been planned or carried out in 6 patients. As PMRDP grows, we aim to expand research collaborations for candidate genes and undiagnosed cases and to implement novel therapies for rare disease patients.

Haplotype-resolved variant detection in medically relevant and paralogous genes using multi-region joint detection with Illumina on-flowcell proximity sequencing

POSTER 307

Shunhua Han[1], Vitor Onuchic1, Stephen Gaffney1, Varun Jain1, Arun Subramaniyan1, Drew Ellershaw1, Sean Humphray1, Mitch Bekritsky1, Simeen Malik1, Alvin Wei Tian Ng3,4, Chong Siao Ting3,4, Inga Nagel2, Inga Vater2, Siew Hong Leong1, Chia Sing Chan1, Wei Wei1, Raden Indah Kendarsari1, Yen Lynn Wong1, Leon Yu1, Jacqui Weir1, Richard Allen1, Kevin Hall1, Joanne Ngeow Yuen Yie3,4, Malte Spielmann2, James Han1

1 Illumina, Inc., San Diego, CA
2 University Hospital Schleswig-Holstein, Kiel & Lübeck, Germany
3 Nanyang Technological University Singapore, Lee Kong Chian School of Medicine, Singapore
4 National Cancer Centre Singapore, Cancer Genetics Service, Singapore

Accurate variant detection in paralogous regions using next-generation sequencing is challenging due to high sequence homology. Current approaches typically require supplementary assays (e.g., MLPA, LR-PCR, or long-read sequencing), increasing cost, complexity, and turnaround time without reliably producing haplotype-resolved calls. Here we present a novel approach by combining the Illumina on-flowcell proximity sequencing, and a multi-region joint detection (MRJD) algorithm, to produce a haplotype-resolved view of complex paralogous regions without prior population haplotype knowledge. This de novo haplotype inference is enabled by linking spatially proximal sequence fragments originating from the same DNA molecule.

We evaluated MRJD with on-flowcell proximity sequencing by comparing haplotype concordance to orthogonal long-read sequencing across clinically significant paralogous loci (SMN1/2, CYP21A2, PMS2, CYP2D6, STRC, OTOA). Our analysis included 14 cell lines with diverse structural complexities. Specifically, 12 cell lines exhibited copy number alterations in SMN1/2, six had recombination in CYP21A2, and 12 had gene conversions in PMS2. MRJD achieved median concordance values of 0.96-0.99 across all tested regions.

We next evaluated this approach using four non-cell-line samples with confirmed Lynch Syndrome diagnoses due to variants in PMS2exon 11-15, including three indels and one complete coding sequence deletion. MRJD resolved all PMS2 and PMS2CL haplotypes, correctly identifying pathogenic variants confirmed by orthogonal tests.

(continue to page 58)

(continued from page 57)

Finally, we applied MRJD to two non-cell-line samples with suspected Congenital Adrenal Hyperplasia, each harboring complex genotypes within the CYP21A2 locus. The first sample showed a carrier genotype based on standard-of-care (SOC) results of the proband and both parents, while the second sample was predicted affected, although MLPA and LR-PCR assays failed to fully resolve the locus. MRJD reconstructed all gene, pseudogene, and fusion copies within the CYP21A2 and CYP21A1P loci, assigned them to specific haplotypes in both cases, and produced haplotype-resolved reconstructions consistent with clinical phenotypes and exceeding the resolution of SOC testing.

MRJD with on-flowcell proximity sequencing represents a significant advancement toward a unified next-generation whole-genome sequencing assay capable of accurately resolving clinically relevant paralogous regions, reducing reliance on supplementary assays.

Harnessing the power of 3D genomics for personalised medicine and drug discovery

POSTER 308

Daniel J. Turner[1], Kishen Chahwala1, Stefan Schoenfelder2, Dietrich Stefan1

1 Enhanced Genomics, Cambridge, United Kingdom
2 Babraham Institute, Cambridge, United Kingdom

Genome-wide association studies (GWAS) have linked thousands of genetic variants to disease, but ~95% of these lie in non-coding regions of the genome. Many variants are located within regulatory elements, which control gene expression in a cell-type-specific manner. These elements can influence genes located hundreds of kilobases or even megabases away, often bypassing several intervening genes. For regulatory elements to affect gene expression, they must be brought into physical proximity with gene promoters through the 3D folding of the genome. This spatial organisation makes it difficult to infer which genes are functionally impacted by disease-associated variants.

By examining 3D genome folding in disease-relevant cell types, Enhanced Genomics identifies which genes are physically and functionally connected to variant-containing regulatory regions. This allows us to uncover disease mechanisms, prioritise genetically supported drug targets, and stratify patients more effectively. In contrast, most genomics approaches treat the genome as a linear sequence and are blind to 3D interactions, limiting their ability to reveal the full landscape of disease biology and to distinguish the most compelling therapeutic targets.

Our pipeline uses promoter-capture Hi-C to map long-range interactions across the entire genome in disease-relevant cell types. We integrate this 3D genome architecture with complementary omics datasets, such as chromatin accessibility, histone modifications and gene expression, into a unified framework we call '3D multiomics'. This integrated view reveals how non-coding variants, regulatory elements, and genes interact within the spatial context of the nucleus.

Our platform reveals that each cell type harbours hundreds of thousands of active regulatory elements, both ubiquitous and cell-type-specific, that engage in long-range interactions with gene promoters to drive context-specific gene expression. By mapping these interactions within the true biological context of the cell, we link disease-associated variants to the genes they influence, enabling the precise and scalable identification of previously unrecognised susceptibility genes and novel therapeutic targets.

We demonstrate a hypothesis-free universal discovery tool that unlocks any genetic disease by identifying its associated genes and relevant cell types, thereby de-risking the drug-discovery process.

Identifying rare disease genes through phenotype embedding from biobank data

POSTER 309

Hongyu Zhao[1]

1 Yale University, New Haven, CT

Following the success of Genome-Wide Association Studies (GWAS) that primarily focus

on common variants, recent years have seen the discoveries of associations between rare variants and phenotypes through the analysis of Whole Exome Sequencing (WES) and/or

Whole Genome Sequencing (WGS) data. The establishment of many large biobanks has

facilitated much of this progress. However, many rare diseases are still underrepresented in large population cohorts, e.g., UK Biobank and the All of Us Research Program, hindering the identifications of pathogenic genes associated with many rare diseases. Few methods take advantage of the extensive Electronic Health Records (EHR) from large biobanks to study rare disease genetics. We hypothesize that by aggregating all phenotype information for each individual, we can better define rare disease status based on the phenotype relationships among different diseases, and better infer pathogenicity of genes on diseases. In this presentation, we will discuss the use of embedding methods, such as Word2Vec, to better embed the diagnosis data based on the phenotype relationships and identify phenotype differences between the potential gene variant carrier group and the non-carrier group.

Illumina on flowcell constellation mapped read technology illuminates dark regions of clinical genomes and facilitates structural variation discovery

POSTER 310

Sergey Batalov[1], Lucitia van der Kraan1, Erica Sanford-Kobayashi1, Jennifer Friedman2, Jerica Lenberg1, Kristen Wigby1, Pascal Grobecker5, Mitch Bekritsky3, Jacqui Weir4, Qing Zhang4, Nafeesa Kahn3, Jon Albay3, Drew Ellershaw4, Ali Crawford3, Matthew N. Bainbridge1

1 Rady Children's Institute for Genomic Medicine, San Diego, CA
2 University of California San Diego and Division of Neurology, Departments of Neurosciences and Pediatrics, Rady Children's Hospital, San Diego, CA
3 Illumina, Inc., San Diego, CA
4 Illumina, Inc., Cambridgeshire, Cambridge, United Kingdom

Whole-genome short read sequencing (SRS) generates high quality data throughout most of the genome but demonstrates limitations in regions of high homology, leading to dark regions of the genome where variants are not detected. Long read sequencing can interrogate these regions and provides more accurate structural variant calls and phasing capabilities. However, single molecule long read sequencing suffers from higher error rates and is more expensive than short read sequencing.

Illumina constellation mapped read technology has all the advantages of short read sequencing (e.g. low indel error rates, ease of use, and low input DNA amounts) while greatly improving the mapping capabilities in regions of high homology. This is accomplished by accounting for the location of each cluster on the flow cell to generate proximity information and virtual long reads which improve read mapping, phasing and structural variant calling

Here we used constellation mapped read technology on 36 clinical samples collected from critically ill children at Rady Children's Hospital. On average, genomes were sequenced to 90-100x depth, had between 85% - 95% of all genes fully phased and an average NG50 phase block of ~1.5Mbp. Constellation mapped read technology recapitulated standard SRS calling 99.3% of all indels and single nucleotide variants ($p<10^{-10}$, t-test). Using 11 clinical samples with known, pathogenic structural and "dark region" variants, we were able to recall all variants including multiple structural variants (deletions/duplications), a stop-loss pathogenic mutation in IKBKG (c.1260G>C p.Ter420TyrextTer27), a pathogenic repeat expansion of DMPK (13/600 repeat copies), and a single base duplication DDX3X (c.7dup). In the remaining 25 samples, we identified several candidate variants that may account for the participant's phenotype: inversion of a single exon in CBLB, intrachromosomal translocations in ADCY2, KCNB2. These copy-neutral structural variants (frequently flanked by repetitive regions) could not be called using SRS and standard variant-caller workflows.

Wide adoption of inexpensive, comprehensive genome sequencing is a critical next step in clinical genomics, helping increase the diagnostic rate of patients, and improved accuracy for structural variants while reducing costs.

Improved RNA-Seq data from archival prostate cancer FFPE tissues using a single-stranded library preparation method without repair

POSTER 311

Varsha Rao[1], Andrew Mishin1, Tobin Groth1, Chris Troll1

1 Claret Bioscience LLC, Scotts Valley, CA

Whole transcriptome analysis using next-generation sequencing is critical to understanding biological process in all organisms, gene expression changes that occur during disease progression and in identification of druggable targets. Standard NGS library preparation methods that involve the ligation of a double-stranded adapters work well with high-quality samples, however they fail with degraded and low quantity samples such as those derived from liquid biopsies and formal-fixed paraffin embedded (FFPE) tissues. Existing methods involve superfluous steps to generate stranded libraries – 2nd strand synthesis, end-repair to facilitate ligation and steps to eliminate or tag the complementary strand to retain directionality information. By contrast, the REALLY™ workflow is a single-stranded approach to library preparation where RNA is fragmented and converted to first-strand cDNA in a single step, followed by single-stranded adapter ligation without 2nd strand synthesis or any additional steps to facilitate ligation. The REALLY workflow is coupled with a streptavidin-bead based depletion step rNONE™ for efficient removal of ribosomal RNA from both human and mouse total RNA. Here we generated RNA-Seq libraries with REALLY-rNONE or a traditional double-stranded method using high quality total RNA (250 ng to 10 ng) or total RNA (10 – 100 ng, RIN < 4) from FFPE tissues (prostate cancer xenografts and tumor or liver tissues). We observed that the REALLY approach captures highly fragmented input and generates high libraries yields with fewer cycles of PCR than double-stranded methods (<5 cycles) which reduces the incidence of PCR errors. Sequencing analyses demonstrated the higher conversion efficiency of the assay and highly complex libraries were generated with inputs as low as 10 ng of total RNA derived from both high-quality total RNA and low-quality degraded RNA derived from FFPE tissues obtained from samples stored for >10 years. Additional transcriptomic analyses show that the method demonstrates uniform coverage across the transcript body (up to ~90%) and high concordance between read counts per million for REALLY-rNONE libraries made with high quality inputs from 10 - 250 ng (R2>0.90) and for FFPE derived RNA (R2 ~0.89) between 10 – 100 ng (traditional RNA-Seq methods failed to generate libraries with 10 ng FFPE RNA, RIN ~3). The workflow builds on Claret Bio's SRSLY technology which can also be used to generate high complexity genomic libraries from degraded FFPE DNA. Together these technologies can be applied in capturing high quality sequencing data from limiting samples settings, often encountered in research and clinical settings.

Interaction of identity and beliefs with genetic literacy using the validated Education and Assessment of Genetic Literacy (EAGL) measure

POSTER 312

Chris Gunter[1], Lily S. Barna1, Gabriela M. Ramírez-Renta1,3

1 National Human Genome Research Institute, Social and Behavioral Research Branch, Engagement Methods Unit, Bethesda, MD
2 National Human Genome Research Institute, Office of the Director, Bethesda, MD
3 University of Colorado Anschutz Medical School, Aurora, CO

Genetic literacy goes beyond knowledge of genetic terms, as it requires sufficient skills and understanding to effectively facilitate health-related decision making and participation in social discussions about genetic issues. Personal identity and beliefs have been shown to affect how individuals interact with new information, but rarely in the context of genetic literacy. We created and disseminated a survey to two separate samples: 2050 members of the US general public and 2023 participants in a large genetic research study. We assessed genetic literacy through three components: subjective knowledge (familiarity), objective knowledge, and knowledge comprehension (skills), making this one of the only large-scale surveys to assess comprehension as a part of genetic literacy.

To better understand how individuals process and retain genetic information, we also asked questions about identity and belief factors. We found that confidence in one's genetic knowledge was the strongest predictor of positive scores in all three components, controlling nearly 25% of the variance in scores, while perceived importance of genetic information had a positive but weaker relationship to scores. This suggests that improving confidence, not just providing knowledge, is an important part of increasing uptake of genetics in various applications. Further, we found that self-described beliefs such as religiosity and political ideology had mixed predictive effects on all three of our genetic literacy subscales. These findings demonstrate the complexity inherent in endeavors to raise genetic literacy in the US population as an example, as well as the importance of context-specific genetics communication.

Finally, to enable wider explorations of genetic literacy as a construct, we validated the Education and Assessment of Genetic Literacy measure (EAGL) in a third survey of 2708 US participants, using Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). After the validation process, we have two measures that are available for wide usage: the multi-faceted EAGL-long, which has previously been tested in thousands of participants, or the validated three-factor EAGL-short. In this large sample, we also found that geographic location in the US measured as metropolitan vs. non-metropolitan status had no significant effects on levels of genetic literacy.

Large language model-based automated phenotyping of Undiagnosed Diseases Network cases

POSTER 401

Inez Oh[1], Aditi Gupta1, Kathleen Sisco2, Erin McRoy2, Dana Kiley2, Jennifer Wambach2, Patricia Dickson2

1 WashU Medicine, Institute for Informatics, Data Science and Biostatistics, St. Louis, MO
2 WashU Medicine, Department of Pediatrics, St. Louis, MO

Evaluating Undiagnosed Diseases Network (UDN) cases currently requires experienced clinicians to manually extract information relevant to diagnosing genetic disorders from patient electronic health records (EHRs), including unstructured clinical narratives. We hypothesized that Large Language Models (LLMs) could be leveraged to increase efficiency of phenotyping UDN patients, due to their ability to handle natural language processing tasks and synthesize information across various contexts.

Using a HIPAA-compliant API endpoint of OpenAI's GPT-4 Omni (GPT-4o) LLM, we developed a pipeline to extract information from clinical narratives on broad domains relevant to diagnosing genetic disorders, such as phenotypes, family history, birth history, laboratory tests, procedures, and referrals to specialists, including dates and results as applicable. Five existing case evaluations were used to validate GPT-4o results. Clinical notes and procedure reports for these five patients were obtained from our institution's EHR database with IRB approval.

We developed zero-shot prompts instructing GPT-4o to extract information on the aforementioned domains without providing specific examples. Hypothesizing that genetics notes would contain most pertinent information, we first provided GPT-4o only with notes written by genetics providers. Subsequently, we prompted GPT-4o to extract the reason for the procedure and any abnormal findings from procedure reports.

Using only genetics notes, GPT-4o retrieved ~50% of the information present in the case evaluations across all patients; for one patient, all information was retrieved. Addition of procedure reports (e.g. MRI, CT, EEG) resulted in retrieval of almost all information except for details not mentioned in the clinical narratives. Error analysis found that information missed by GPT-4o was not present in the narratives, either because additional detail was not documented (e.g. variants of uncertain significance), or reflecting a normal lag in the clinical workflow (e.g. between ordering and resulting of a test).

Our results show that LLMs can be feasibly leveraged for patient phenotyping, but completeness depends on providing informative sources to the model. Further work in developing this pipeline entails including notes from additional specialties and integrating structured data such as lab results to ensure that up to date information is available for UDN case evaluations.

Launch of the Genomics-enabled Learning Health Systems (gLHS) network: improving approaches for integration of genomic medicine in clinical care

POSTER 402

Jodell E Linder[1], Carolyn Audet2, Rinad S. Beidas3, Alexander Bick4, Adam H. Buchanan5, Andrea Burnett-Hartman6, Paul R. Dexter7, Patricia D. Franklin3, Lucy A. Godley8, Cheryl L. Gatto1, Megan He1, Kensaku Kawamoto9, Sunil Kripalani4, Alanna Kulchak Rahm10, Teri A. Manolio10, Cara Z. McCormick5, Elizabeth M. McNally8, Lori A. Orlando11, Dan Roden12, Robb Rowley10, Maren Scheuner13,14, Todd C. Skaar15, Martin Tristani-Firouzi16, Jason L. Vassy13,17, Deepak Voora13,18, Mark Yandell19, gLHS Network10, Josh F. Peterson20

1 Vanderbilt University Medical Center, Vanderbilt Institute for Clinical and Translational Research, Nashville, TN
2 Vanderbilt University Medical Center, Health Policy and the Institute for Global Health, Nashville, TN
3 Northwestern University Feinberg School of Medicine, Department of Medical Social Sciences, Chicago, IL
4 Vanderbilt University Medical Center, Department of Medicine, Nashville, TN
5 Geisinger Health, Department of Genomic Health, Danville, PA
6 National Institutes of Health, National Cancer Institute, Bethesda, MD
7 Indiana University School of Medicine, Department of Clinical Medicine, Indianapolis, IN
8 Northwestern University Feinberg School of Medicine, Department of Medicine, Chicago, IL
9 University of Utah Health, Department of Biomedical Informatics, Salt Lake City, UT
10 National Institutes of Health, National Human Genome Research Institute, Division of Genomic Medicine, Bethesda, MD
11 Wake Forest University, Department of General Internal Medicine, Winston-Salem, NC
12 Vanderbilt University Medical Center, Department of Medicine, Pharmacology, and Biomedical Informatics Nashville, TN
13 United States Department of Veterans Affairs, Washington, DC
14 University of California, San Francisco, Department of Medicine and Pediatrics, San Francisco, CA
15 Indiana University School of Medicine, Department of Medicine, Indianapolis , IN
16 University of Utah Health, Department of Pediatrics, Salt Lake City, UT
17 Harvard Medical School, Department of Medicine, Boston, MA
18 Duke University, Department of Medicine, Durham, NC
19 University of Utah Health, Department of Human Genetics, Salt Lake City, UT
20 Vanderbilt University Medical Center, Department Biomedical Informatics, Nashville, TN

The translation of genomic discovery to clinical practice has yet to reach its full potential in guiding healthcare and improving health outcomes.

(continue to page 66)

(continued from page 65)

Barriers include the limited scope of previous implementation efforts, lagging payor reimbursement policies, lack of interoperability of genomic data, limited clinician familiarity with genetic testing protocols, limited demonstration of clinical utility, and non-scalable approaches to returning genetic data. To address these issues, the Genomics-enabled Learning Health System (gLHS) network was created to facilitate adoption of recommended genomic medicine practices across varied health systems and clinical settings. This abstract describes the organization of the network and development of tools enabling the selection and implementation of the network projects. Funded by the NIH in September 2024, the network is composed of 6 clinical sites, 8 community partners, and a coordinating center (CC). The aim of gLHS is to identify and improve approaches for clinical integration of genomic medicine strategies in a learning health system cycle of implementation, assessment, refinement, and re-implementation. The network will develop and implement 2-4 genomic intervention projects across the network deployed through implementation cycles with outcome measurements. During the first 3 months, the network developed a framework focused on clinical utility and implementation gaps, financial implications, health care access, impact, and feasibility to weigh the strengths of potential implementation projects. Three action groups focused on implementation science, clinical science and health systems, and electronic health record tools and data analysis drive the network. The CC developed tools and frameworks to facilitate the selection of the first project focused on the adoption of guideline-concordant germline testing. Germline testing for hereditary cancer syndromes informs therapies, surveillance, and cascade testing for family members. Given the current know-do gap and underutilization of testing, the project is focused on improving the implementation of germline testing for hereditary cancer syndromes among cancer patients. The goal is to increase the number of guideline-concordant test orders placed by clinicians (i.e. mainstreaming of genetic testing). The network will implement the project in the fall of 2025 and continue development of a second project in 2026.

Learning a transcriptomic landscape of endometriosis using a single-cell foundation model

POSTER 403

Umair Khan[1,2], Ana Almonte Loya1,2,3, Wanxin Wang4, Sahar Houshdaran4, Júlia Vallvé-Juanico4, Divya Kushnoor3, Carlos Simón5,6, Kim Chi Vo4, Juan C. Irwin4, Alexis J. Combes3, Gabriela K. Fragiadakis3, Christina V. Theodoris7,8, Michael J. Keiser9, Linda C. Giudice4, Marina Sirota1,8

1 Bakar Computational Health Sciences Institute, University of California, San Francisco, San Francisco, CA, USA
2 Biological and Medical Informatics Graduate Program, University of California, San Francisco, San Francisco, CA, USA
3 CoLabs, University of California, San Francisco, San Francisco, CA, USA
4 Center for Reproductive Sciences, Department of Obstetrics, Gynecology and Reproductive Sciences, University of California, San Francisco, San Francisco, CA, USA
5 INCLIVA Biomedical Research Institute, Valencia, Spain
6 Department of Pediatrics, Obstetrics and Gynecology, School of Medicine, University of Valencia, Valencia, Spain
7 Gladstone Institutes, San Francisco, CA, USA
8 Department of Pediatrics, University of California, San Francisco, San Francisco, CA, USA
9 Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA, USA

Endometriosis is scored using American Society for Reproductive Medicine guidelines (rASRM) into one of four stages based on lesion location and severity. It remains unclear how transcriptomic features of endometriosis relate to this clinical score. Emerging single-cell data, in combination with novel machine learning models that capture broad biological knowledge, have the potential to address this gap. We applied Geneformer, a foundation model pretrained on millions of cells across diverse tissue contexts, to a dataset of over three hundred thousand single-cell transcriptomes from endometrial tissue in 27 cases and 19 controls. We find that Geneformer recapitulates biological structure in this dataset without any further training. We then fine-tuned Geneformer to learn a landscape of disease stage by predicting rASRM score (or absence of disease) given a transcriptome. We find that key genes driving the structure of this learned latent space (including PAEP and CXCL10) align with prior knowledge. With further optimization, this approach can shed light on whether transcriptomic markers have any diagnostic or therapeutic potential by learning a landscape of endometriosis stages compared to control cells.

Long read methylation analysis of FSHD cell lines reveal differential methylation between haplotypes at DUX4 targets and hypomethylation at complex repeat regions

POSTER 404

Jasmine Sakr[1], Xiangduo Kong2, Kyoko Yokomori2, Ali Mortazavi1

1 University of California, Irvine, Department of Developmental and Cell Biology and Center for Complex Biological Systems, Irvine, CA, United States
2 University of California, Irvine, Department of Biological Chemistry, Irvine, CA, United States

Facioscapulohumeral muscular dystrophy (FSHD) is the third most prevalent muscular dystrophy and is linked to a mono-allelic contraction of primate-specific D4Z4 macrosatellite repeats on the disease-permissive chromosome 4q (4qA haplotype) with additional mutations of a chromatin regulator SMCHD1 acting as a disease modifier. D4Z4 repeats form heterochromatin, and its disruption (loss of DNA methylation) and resulting abnormal derepression of an embryonic transcription factor DUX4 encoded in the D4Z4 repeat are the hallmark of FSHD. However, defining the precise effect of FSHD mutations has been difficult. Each unit of the D4Z4 repeat array is 3.3 kb and the array ranges from one to over one hundred units, making this locus difficult to characterize with short-reads alone. We performed Nanopore sequencing to characterize global and D4Z4-specific changes in DNA methylation using our CRISPR-engineered human skeletal myoblast lines carrying FSHD mutations (D4Z4 contraction and SMCHD1 mutation) compared to the isogenic parental healthy control line. Nanopore reads longer than 60 kb allowed us to define the entire unedited control and contracted mutant D4Z4 arrays. Long-reads distinguish differential methylation patterns at the disease locus on chromosome 4qA from those at a nearly identical nonpathogenic D4Z4 repeat arrays on chromosome 10 and disease non-permissive 4qB allele. We observe global hypomethylation in mutant cells, most notably at satellite repeats. Allele-specific analyses of 5mC and 5hmC in myoblasts and differentiated myotubes reveal allelic differences in methylation patterns at multiple DUX4 target promoters in mutant cells. Taken together, our results indicate significant impact of FSHD mutations on D4Z4 allele, DUX4 targets and repeat regions in the genome, which may be collectively contributing to the FSHD pathogenesis.

Multiomics investigation of Ring 14

POSTER 405

Emily Bonkowski[1], Christy LaFlamme1, Brett Brown1, Juana Manuel2, Edith Almanza Fuerte1, Soham Sengupta1, Wani Arjumand3, Alfonso Lavado1, Jasmine Plummer1,3, Heather C. Mefford1

1 St. Jude Children's Research Hospital, Department of Cell and Molecular Biology, Center for Pediatric Neurological Disease Research, Memphis, TN, USA
2 University of Tennessee Health Science Center Integrated Biomedical Sciences; Department of Genetics, Genomics, and Informatics; Memphis, TN, USA
3 St. Jude Children's Research Hospital, Department of Developmental Neurobiology, Memphis, TN, USA

Background: Ring 14 [r(14)] syndrome is a rare chromosomal disorder characterized by the formation of a ring structure on chr14. There are fewer than 100 known cases of r(14) syndrome, which manifests in life-limiting symptoms, including intractable epilepsy, hypotonia, and intellectual disability. Ring chromosomes form due to breakage and fusion of chromosome ends, often resulting in loss of genomic material through terminal or interstitial deletions containing protein-coding genes. Ring chromosomes are hypothesized to alter epigenetic landscapes, potentially affecting gene regulation. Few studies have systematically investigated or modeled ring chromosomes due to their genomic instability. Mosaicism has been observed in cellular models and has been hypothesized in affected individuals with r(14). The pathophysiological mechanisms of r(14) syndrome remain poorly understood. Methods: We generated multiomic data using DNA and RNA from peripheral blood for twelve r(14) cases and post-mortem brain tissue from one r(14) case and a matched control. Since the resolution of standard diagnostic karyotyping is ⩾5Mb and short-read sequencing cannot map highly repetitive sequences in acrocentric arms, we used Oxford nanopore long-read sequencing (LRS) aligned to the telomere-to-telomere (T2T) reference to map precise r(14) breakpoints. We investigated differential methylation for nine r(14) probands compared to 15 controls. Using RNA-seq, we evaluated differential expression of seven r(14) probands compared to 19 controls to identify candidate genes related to phenotype. We performed bulk RNA-seq, spatial transcriptomics, LRS, and DNA methylation studies using post-mortem brain tissue. Results: LRS successfully resolved r(14) with high resolution of breakpoints. The breakpoints differ in 7/12 cases analyzed to date; two cases harbor r(14) breakpoints that disrupt a known epilepsy gene PACS2, suggesting convergent mechanisms. LRS of post-mortem brain tissue across 4 cortical regions indicates that the ring chromosome is constitutively heterozygous in all regions tested without evidence of somatic mosaicism in the brain. Epigenomic and transcriptomic analyses are ongoing. Conclusion: Multiomic investigations of r(14), including in disease-relevant tissue, offer insights into disease mechanisms of r(14) applicable to other ring chromosome disorders and may inform future therapeutic development. Additionally, the ability of LRS to detect ring structures and resolve their breakpoints increases the diagnostic utility of LRS.

NFIX missense variants in key DNA-binding residues lead to a muscle phenotype distinct from Malan or Marshall-Smith syndromes

POSTER 406

Benjamin T. Cocanougher[1,2], Olivia M. D'Annibale1,2, Kaitlynn Stowers1,2, Christal G. Delagrammatikas3

1 Cincinnati Children's Hospital, Division of Human Genetics, Cincinnati, OH
2 University of Cincinnati College of Medicine, Department of Pediatrics, Cincinnati, OH
3 Malan Syndrome Foundation, Old Bridge, NJ

Genetic variation in the transcription factor NFIX is known to cause two distinct allelic disorders: (1) Marshall-Smith syndrome and (2) Malan syndrome. Marshall-Smith syndrome is caused by heterozygous truncating variants in exons 6-11, leading to a toxic gain-of-function protein that can bind DNA, but is unable to coordinate protein-protein interactions. Most individuals with Marshall-Smith have severe dysostosis and respiratory insufficiency, and many do not survive past infancy.

Malan syndrome results from heterozygous NFIX gene deletions or loss-of-function mutations, predominantly affecting exon 2, which cause haploinsufficiency through reduced NFIX protein expression. Patients with Malan syndrome have a milder phenotype than Marshall-Smith syndrome. Malan syndrome is characterized by skeletal overgrowth and moderate-to-severe intellectual disability. From cases reported in the literature to date, lifespan in Malan syndrome does not appear to be significantly shortened.

Here we describe a novel NFIX-related disorder in a cohort of patients that is most consistent with Malan syndrome in infancy and early childhood; unlike Malan syndrome, however, it is characterized by progressive muscle loss and scoliosis in adolescence. This rapidly progressive scoliosis, paired with muscle weakness, was lethal in at least two cases. The patients in this cohort share de novo missense variants in conserved DNA-binding residues of the beta-hairpin loop within exon 2. Our working hypothesis is that these missense variants confer a toxic gain-of-function protein. This mechanism is distinct from the haploinsufficiency of Malan syndrome and the toxic gain-of-function affecting protein-protein interactions of Marshall-Smith syndrome. Ongoing studies include CHIP-seq in patient-derived samples to test if these missense variants alter NFIX binding, serum biomarker investigation, and exploration of therapeutic strategies. These findings expand

the phenotypic spectrum of NFIX mutations and highlight the potential for precision therapeutic interventions in rare myopathic disorders.

Novel Host Stool mRNA Extraction Method Improves Sensitivity of Detection for Cancerous and Precancerous Lesions

POSTER 407

Jakob Kirchner, Loren Hanson, Wenying Pan

El Capitan Biosciences, Pleasanton, CA, 94566

Colorectal cancer (CRC) is the second leading cause of cancer-related deaths world-wide. Early detection is crucial, but the invasiveness and resource demands of colonoscopy limit its use. Non-invasive alternatives like fecal immunochemical tests (FIT) and multitarget DNA assays exist but have varying accuracy, particularly in detecting advanced precancerous lesions (APL) and early-stage CRC. Identification of host nucleic acid biomarkers including host mRNA is a promising approach to build improved CRC screening assays.

Most existing methods for nucleic acid extraction from stool were designed for isolation of metagenomic nucleic acids. Most cannot easily be adapted for larger stool input and do not fully remove PCR inhibitors. We have developed a novel extraction method for isolation of host mRNA from stool that is quick and scalable, isolates eukaryotic mRNA directly while removing inhibitors, and gives much improved sensitivity of detection and thus better diagnostic performance than a "gold standard" phenol chloroform extraction method.

The method uses hybridization capture of mRNA under high denaturing conditions by adding locked nucleic acid bases (LNAs) to the capture probe. LNA bases have an "extra" bridge which locks the base in the 3' endo confirmation, improving stability of base-pairing under certain conditions. We take advantage of these properties to directly isolate host mRNA from crude stool lysate under the denaturing conditions found during cell lysis.

We tested our method on clinical stool samples comprising 29 CRC, 19 APL and 26 normal samples, measuring a panel of 11 mRNA biomarkers and a normalization gene. The improved sensitivity of detecting the host mRNA transcripts translated to improved diagnostic performance. The sensitivity for colorectal cancer detection was 96%, and the APL sensitivity was 52% at a specificity of 93%. Compared to the phenyl chloroform-based method with a sensitivity of detection for CRC lesions of 88% and a sensitivity of APL detection of 34% at 92% specificity. Our results demonstrate that improving the sensitivity of detection for certain mRNA biomarkers, can substantially improve the ability to detect malignant lesions without compromising specificity.

Partitioning of circulating nucleic acids from plasma for liquid biopsy diagnostics

POSTER 408

Austin P. So[1], Alexander J. So1

1 Clarica Genomics, Inc., Livermore, CA

Circulating nucleic acids (CNAs) in bodily fluids, and in particular blood, has garnered significant interest as a repository of disease biomarkers. Genomic assays targeting CNAs are however limited by upstream pre-analytical variables: although plasma DNA and RNA can arise from diverse tissues, mitochondria, microbes, and transplanted organs and can exist in multiple molecular states, current workflows homogenize this complexity prior to analysis. Not only does this lose contextual information, but low-abundance targets such as rare somatic variants, pathogen genomes, or graft-derived fragments, become diluted and can only be detected by deploying ultra-deep sequencing.

To address this, we describe DEPLETE, a reagent-based platform that enables the segregation of CNAs directly from plasma according to its state in vivo. Based on the ability form nucleosome arrays in vitro, DEPLETE enables the non-destructive partitioning of CNAs into three native fractions, namely a mitochondrial-rich, a nucleosomal-rich, and a soluble/microbial-rich fraction, as determined by qPCR. Regions in KRAS and TP53 gene loci known to be inaccessible in peripheral blood mononuclear cells (PBMCs) according to ATAC-seq data are also depleted in the soluble/microbial-rich fraction. Low-depth WGS across all fractions demonstrate a bias in sequence content that confirms the ability to segregate CNAs into functionally relevant species. Partitioning of CNAs thus presents a unique opportunity to expand the diagnostic yield of liquid biopsies.

Pathogenic transposon insertions and individualized intravitreal antisense oligonucleotide therapy for FLVCR1-related retinitis pigmentosa

POSTER 409

Boxun Zhao[1,*], Emily A. McCourt2,3,*, Austin Larson2,3, Joseph Ochaba4, Scott Henry4, Anna Howell2, Timothy C. Kenny5, Victoria Suslovitch1, Diana H. Chin1, Lynn W. Bush1, Karl R. Koehler1, Kıvanç Birsoy5, Deborah Chiabrando6, Laurence Mignon7, Julie Douville7, M. Natalia Vergara2, Scott T. Demarest2,3, C. Frank Bennett4, Eunjung Alice Lee1,#, Timothy W. Yu1,#

1 Boston Children's Hospital, Boston, MA
2 University of Colorado, Aurora, CO
3 Children's Hospital Colorado, Aurora, CO
4 Ionis Pharmaceuticals, Inc., Carlsbad, CA
5 Rockefeller University, New York, NY
6 University of Turin, Torino, TO, Italy
7 n-Lorem Foundation, Carlsbad, CA

Transposon insertions represent a significant yet understudied source of disease-causing variants. To understand their contribution to the genetic architecture of Mendelian disorders, we analyzed genomic sequencing data from patient cohorts and identified a diverse collection of pathogenic insertions (L1Hs, Alu, SVA, and pseudogenes), including inherited, de novo, and mosaic events. Notably, many of these insertions disrupt RNA splicing, presenting new therapeutic opportunities. Antisense oligonucleotides (ASOs)—short, chemically modified RNAs—offer a programmable approach to correct mis-splicing and have been used to develop patient-customized therapies for several severe neurological conditions.

Posterior column ataxia with retinitis pigmentosa (PCARP) is an autosomal recessive disease caused by biallelic mutations in the FLVCR1 gene, which encodes a choline transporter. Individuals with PCARP suffer progressive blindness and sensory neuropathy. Here, we report the development of a patient-customized ASO targeting a mis-splicing event in a child with PCARP. Genome sequencing revealed a ~3 kb ISG20L2 pseudogene insertion in intron 8 of FLVCR1. RNA sequencing confirmed that the pseudogene insertion activated a pair of novel splice sites, creating a truncating pseudoexon that led to the degradation of transcripts by nonsense-mediated decay. We designed and tested a splice-switching ASO that restored normal splicing and protein expression in patient-derived cell and organoid models. Supported by robust preclinical efficacy and safety data, this investigational ASO was advanced to an investigator-initiated trial and administered intravitreally to the affected child. After 1.5 years of treatment, the ASO treatment has been well tolerated, with no dose-limiting toxicities, and promising early signs of clinical benefit.

Our study highlights retroelement insertions as an important class of causal, therapeutically targetable variants in genetic diseases. We show that developing new therapies for patients with rare conditions requires the integration of comprehensive genomic analysis, a clear understanding of the pathogenic mutation, rational drug design and development, and careful interventional considerations. In summary, this work illustrates the feasibility of individualized ASO development for rare inherited retinal conditions and highlights the expanding clinical potential of personalized RNA-targeted therapeutics beyond the brain and spinal cord.

Personalized splice-modulating antisense oligonucleotide therapy for PEX1-related Zellweger spectrum disorder (ZSD)

POSTER 410

Didem Demirbas[1], Robert Thompson1,2, Claudia Lentucci1, Kaitlyn Phillips1, Aubrie Soucy1, Emma Sherrill1, Wei Cui3, Erminia Di Pietro3, Lynn Bush1, Margaret Meserve1, Minh Nguyen1, Boxun Zhao1, April Hu1, David Sweetser2, Florian Eichler2, Nancy Braverman3, Timothy W. Yu1

1 Boston Children's Hospital, Boston, MA
2 Massachusetts General Brigham, Boston, MA
3 McGill University, Montreal, Canada

Recent advances have opened the door to highly individualized genetic therapies, offering interventional opportunities in severe, otherwise untreatable Mendelian disorders. Here, we developed a genetically targeted, patient-customized antisense oligonucleotide (ASO) therapy for PEX1-associated Zellweger Spectrum Disorder. ZSD is a genetic disorder caused by autosomal recessive variants in genes encoding for peroxins, proteins that are critical for the biogenesis and functioning of peroxisomes. These mutations disrupt peroxisomal function, resulting in abnormal clearance of very long chain fatty acids and other metabolic disturbances that lead to progressive, debilitating neurological and multi-organ system disease. Death frequently occurs in infancy or childhood and there are no treatments currently available.

We identified a 2-year-old male with ZSD caused by compound heterozygous variants in PEX1 with moderate to severe disease burden which included neurologic, respiratory, GI, and liver complications. This individual has a pathogenic, deep intronic PEX1 c.1359+601A>G variant that activates a 169 bp pseudoexon in intron 6. RT-PCR and RNA-sequencing of patient fibroblasts confirmed pseudoexon activation resulting in loss of PEX1 transcript and protein. Patient fibroblasts also exhibited impaired peroxisomal import as well as biochemical abnormalities (elevated C26:0 LPC, decreased plasmalogen levels) pathognomonic for ZSD.

Given the progressive nature of his symptoms and the lack of existing treatments, along with the ASO amenability of the deep intronic variant, this case was deemed a strong candidate for a personalized ASO therapy. We designed and tested ASOs to block the cryptic splice donor site. Administration of these ASOs to patient fibroblasts and iPSCs increased the expression of properly spliced PEX1 mRNA by two to three-fold in a dose dependent manner. ASO treatment also increased PEX1 protein levels. Finally, ASO treatment rescued defective peroxisomal import, and corrected ZSD-associated biochemical abnormalities (reduced abnormal elevation of C26:0 LPC, and normalized plasmalogen levels) in fibroblasts– thus providing clinically relevant biomarkers for a potential clinical N=1 trial. Unfortunately, before such a trial could be initiated, the index patient passed away due to progressive respiratory compromise. Nonetheless, these results provide a pilot template for potential future personalized ASO therapies for peroxisomal disorders.

Poincaré plots for precision health: a novel visualization tool for interpreting continuous glucose monitoring (CGM) data as part of a streamlined analysis pipeline for clinical and research use

POSTER 411

Kaja Falkenhain[1], Diana M. Thomas2, G. Jake LaPorte2, Corby K. Martin1, Leanne M. Redman1

1 Pennington Biomedical Research Center, Baton Rouge, LA
2 US Military Academy, Department of Mathematical Sciences, West Point, NY

Continuous glucose monitors (CGM) are increasingly being employed to gain insight into glycemic dynamics. Challenges with the interpretation of these data exist due to inconsistencies with simplifying the data from high-frequency time-series measures to a set of comparable glycemic measures. A Poincaré plot is a novel analytic and visualization tool in this context; for CGM data, glucose concentration is plotted against its successive concentration, forming an elliptical, two-dimensional phase portrait. Summative metrics can then be derived to offer a visual and computational supplement to traditional metrics. Using a streamlined, open-source analysis pipeline (GV-Calc), we computed and compared Poincaré-derived numerical axis metrics (i.e., the major and minor axis of the ellipse) with internationally accepted standard CGM metrics in data from N=50 healthy individuals. The minor axis showed the strongest relationships with measures of glycemic variability (mean amplitude: r=0.93, P<0.001; standard deviation [SD]: 0.51, P<0.001; coefficient of variation [CV]: 0.48, P<0.001). The major axis was correlated with both measures of glycemic variability (SD: r=0.99, P<0.001; CV: 0.80, P<0.001; interquartile range: r=0.76; P<0.001; mean amplitude: r=0.44, P=0.001) and control (time in normal range: r=-0.60, P<0.001; mean: r=0.41, P=0.003). Furthermore, the relationship between glycated hemoglobin (HbA1c) – a clinical chemistry measure of long-term glycemic control – with the major axis (r=0.31, 95% CI: 0.02; 0.55, P=0.03) was comparable to that of HbA1c with mean glucose (r=0.30, 95% CI: 0.0; 0.54, P=0.05). These readily computed Poincaré plots and curve/shape parameters appear to yield information about both glycemic control and variability. The ellipse's major axis captures long-term glycemic control, whereas its minor axis reflects short-term glycemic volatility. Future work using CGM data collected in large-scale trials including those from the Nutrition for Precision Health study is needed to confirm the clinical utility of metrics derived from Poincaré plots as a phenotyping modality in precision health (including whether these metrics are able to provide additional value beyond existing parameters) and to inform standardized threshold(s) to differentiate between healthy populations and those at risk for or living with dysregulated glucose metabolism (e.g., type 2 diabetes).

Practical approaches to integrate genomic and transcriptomic analyses to improve the diagnostic yield in a heterogenous cohort of rare neuromuscular disorders

POSTER 501

Vijay Ganesh[1-4], Jialan Ma1, Ben Weisburd1, Stephanie DiTroia1, Anne O'Donnell-Luria1,3,4

1 Broad Institute of MIT and Harvard, Center for Mendelian Genomics, Program in Medical and Population Genetics, Cambridge, MA
2 Brigham and Women's Hospital, Department of Neurology, Boston, MA
3 Boston Children's Hospital, Division of Genetics and Genomics, Boston, MA
4 Harvard Medical School, Boston, MA

Over half of individuals with Mendelian neuromuscular disorders lack a specific molecular diagnosis. Although the widespread adoption of exome sequencing—and the emerging application of clinical genome sequencing—has incrementally improved diagnostic sensitivity in these disorders, the ability to detect rare variants continues to outpace our ability to interpret their functional consequences. Transcriptome sequencing (RNA-seq) has emerged as a method to improve variant resolution and diagnostic yield in rare disease. Prior work at the Broad Center for Mendelian Genomics was the first to demonstrate the utility of RNA-seq for genetic diagnosis in Mendelian neuromuscular disorders specifically. We report our retrospective experience from an expanded cohort of 333 individuals with suspected Mendelian neuromuscular disorders who underwent genome sequencing from whole blood and RNA sequencing from clinically accessible tissues (193 muscle, 140 cultured fibroblasts), analyzed in three stages. In the first stage, comprehensive "genome-first" analysis—including structural variant (SV), mitochondrial DNA, and short tandem repeat (STR) expansion calling—identified a causal variant in 27% of individuals. In the second stage, locus-specific visualization of RNA-seq data provided functional validation of noncoding variants of uncertain significance with predicted splicing disruption in an additional 14%. In the third stage, "transcriptome-first" approaches to RNA-seq, using a combination of expression outlier detection methods (OUTRIDER, FRASER), identified transcriptional patterns in 6% of individuals that led to the identification of the causal gene or suggested new gene-disease associations. Altogether, our cohort analysis underscores the essential role of genome and transcriptome sequencing in improving the low diagnostic rate in Mendelian neuromuscular disorders, and we propose a practical algorithm for implementing such analyses in clinical diagnostic workflows.

Shariant: A National Platform for Sharing Genomic Variant Classifications in Australia

POSTER 502

David Lawrence[1], James Andrews1,2, Thomas Litster1, Emma Tudini2,3, Hamish Scott1,2,4,5, Amanda Spurdle3

1 Centre for Cancer Biology, SA Pathology and University of South Australia, Australian Cancer Research Foundation Cancer Genomics Facility, Adelaide, South Australia, Australia
2 Australian Genomics, Melbourne, Victoria, Australia
3 QIMR Berghofer, Population Health, Brisbane, Queensland, Australia
4 SA Pathology, Genetics and Molecular Pathology, Adelaide, South Australia, Australia
5 The University of Adelaide, Adelaide Medical School, Adelaide, South Australia, Australia

Shariant (https://shariant.org.au) is a platform that enables Australian clinical genetic testing labs to share germline and somatic variant classifications and evidence. Operating as a secure, private environment, it enables labs to identify and reconcile interpretation differences before releasing to public resources such as ClinVar, ensuring higher confidence and consistency.

Shariant supports fully automated API-based data exchange with labs' curation systems. Ingested data undergoes cleaning, variant resolution, and liftover to other genome builds (GRCh37/GRCh38). Issues are flagged for manual review. For historical HGVS resolution, we processed 70 GTF files from RefSeq and Ensembl to build a dataset of 1,398,973 transcripts, released as the open-source project "cdot"

We recently added support for some types of structural variants, which introduced challenges in normalization, overlaps, liftover, and limited toolchain support.

Our team works with labs to overcome data integration challenges of heterogeneous systems and map them to a common format. Shariant stores discipline-specific classification criteria and an evolving set of supporting evidence in JSON, enabling complete capture of structured and unstructured evidence, with validation and warnings

Classification discordances are detected automatically, triggering a triage workflow between affected labs. Detailed recording of evidence against classification criteria allows for side-by-side comparison of classifications, contributing to faster resolution.

The web interface supports export of classification data for integration into lab systems and allele and gene browsing in any build, with a flexible search engine that handles 12 query formats, and maps non-standard HGVS strings (e.g. "CFTR:c.1438G>T") using MANE transcripts.

To date, >40,000 records (>29,000 unique alleles) have been shared across 18 Australian labs contributing to Shariant. Among variants with multiple contributors, 13.6% were discordant, and 50% of medically significant cases have since been resolved. Shariant has enabled submission of over 13,000 records to ClinVar, a 3000% increase in Australian contributions.

Shariant demonstrates that an automated classification sharing platform integrates into busy lab workflows to sustain national sharing, accelerates resolution of classification discordance by focussing inter-lab discussion and contributes to more consistent and accurate classification and reporting.

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

POSTER 503

Neil Slaven[1], Sarah Morton2, Yuri Kim2, Michal Kosicki1, Laura E Cook1, Momoe Kato1, Ingrid Plajzer-Frick1, Christine Seidman2, Jonathan Seidman2, Len Pennacchio1, Axel Visel1

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) is the most common birth defect worldwide and affects millions of individuals in the US alone. Despite significant efforts to identify the genetic causes of CHD, coding variants are estimated to explain only ~30% of cases. Seminal studies have shown that non-coding variants in regulatory elements, such as transcriptional enhancers, can contribute to congenital disease, suggesting that unresolved cases may be explained by such variants. However, the global contribution of heart enhancer variants to CHD remains unclear. To address this gap, we prioritized non-coding variants from a unique CHD patient cohort and characterized their functional impact in vivo. Whole-genome sequencing was performed by the Pediatric Cardiac Genomics Consortium on 1,831 trios with severe cardiac defects and no evidence of pathogenic coding mutations. This effort identified over 75,000 de novo and more than 1 million rare non-coding variants. Intersecting these with a predicted heart enhancer catalog yielded ~1,000 de novo and ~220,000 rare candidate variants. We further prioritized variants using multiple lines of evidence, including H3K27Ac enrichment, chromatin accessibility, evolutionary conservation, and predicted disruption of heart transcription factor binding sites. In vivo testing of wild-type and variant enhancer sequences was performed using a reporter knock-in mouse assay (enSERT). To date, we have identified 20 novel developmental heart enhancers, with 7 (35%) showing altered activity upon introduction of patient-derived variants. In two cases, multiple variants from different patients disrupted the same enhancer. These results indicate that non-coding variants can frequently impact heart enhancer activity (~35%), although their role in CHD pathogenesis remains to be fully understood. We are currently generating enhancer knockout and knock-in mouse models to assess the developmental consequences of these variants, offering insight into the frequency and significance of heart enhancer mutations in CHD.

Rapid whole genome sequencing for unexplained cardiac arrest in pediatric patients

POSTER 504

Erica Sanford Kobayashi[1,2], Ofelia Vargas-Shiraishi2, Jason Knight2, Emily Eckhardt3, Monia Hammer1, Lucia Guidugli1, Stephen F. Kingsmore1, Matthew Bainbridge1, Kristen Wigby1

1 Rady Children's Institute for Genomic Medicine, San Diego, CA
2 Children's Hospital Orange County, Department of Pediatrics, Orange, CA
3 Rutgers University, School of Nursing, Newark, NJ

Rapid whole genome sequencing (rWGS) is transforming pediatric intensive care (ICU) medicine by providing fast, accurate diagnosis of genetic disorders in critically ill infants and children, thereby facilitating timely initiation of disease-specific therapies, which can be crucial for mitigating morbidity and mortality. Changes in clinical management for diagnosed patients have included starting/stopping medications, dietary changes, avoidance/indication for a surgery or invasive procedure, and expedient evaluation for organ transplantation. Though clinical presentations leading to rWGS in ICUs are heterogeneous, some specific subgroups of children have emerged for which the available utility evidence is compelling, one of which is unexplained cardiac arrest (UCA). UCA is an uncommon but devastating event with high morbidity and mortality. Life-threatening channelopathies such as long QT syndrome and catecholaminergic polymorphic ventricular tachycardia have been found in up to 54% of pediatric UCA. Despite this, genetic testing has not been recommended as standard of care for UCA, and historically, even when undertaken, has been limited to a short list of cardiac genes rather than WGS. Furthermore, genetic testing has hitherto been postmortem or remote in time to UCA and primarily focused on identification of at-risk family members. With increasingly reimbursed rWGS, the opportunity to make a diagnosis while children are still receiving inpatient care has emerged, thereby offering the potential to directly impact the clinical course of patients with UCA. Here, we undertook a retrospective evaluation of a cohort of critically ill pediatric patients who were admitted to ICU at a single tertiary children's hospital and received rWGS. We identified 14 patients, aged 0 days to 13 years, whose indication for rWGS was UCA. A molecular diagnosis was made in 9 of the 14 patients (64%), representing 9 unique molecular disorders. Seventy-eight percent of children had a resultant change in management. In combination with the other available evidence, this lends support for application of rWGS for expedited diagnosis in UCA in pediatric ICU patients. We have proposed a hospital-specific clinical practice guideline recommending rWGS as standard in cases of UCA to optimize care for these critically ill patients.

Rare disease diagnosis through genome sequencing in large Middle Eastern families in Qatar

POSTER 505

Aljazi Al-Maraghi[1], Waleed Aamer1, Elbay Aliyev1, Rulan Shaath1,2, Najeeb Syed1, Amal Ibrahim2,3, Rafah Mackeh1, Alya A. Al-Kurbi1, Geethanjali Devadoss Gandhi1, Maria Farag4, Fazulur Rehaman Vempall5, Fatima Al-Khouri4, Sura Hussein1, Shima Salim1, Reem Al-Karbi1, Jyothi Lakshmi1, Sasirekha Palaniswamy1, Saja Baheer4, Sujitha Padmajeya1, Omayma Al-Saei1, Abubakr Imam6, Ibrahim Shatat6, Ahmed Kaddoura6, Donald Love7, Marios Kambouris7, Noora Shahbeck4, Dhullipala Anand8, Ibrahim Janahi9, Ziyad Hijazi10, Khalid Omer11, Mahmoud Fawzi11, Shiga Rappai12, Ghassan Mohamadsalih12 , Jehan Al Rayahi13, Chiara Cugno1, Bernice Lo1, Ruba Benini11, Khalid Hussain12, Rehab Al-Saleh4, Sonia Davila1, Tawfeg Ben-Omran4, Ammira S.A. Akil1, Khalid A. Fakhro1,2,14*

1 Sidra Medicine, Department of Human Genetics, Doha, Qatar
2 Hamad Bin Khalifa University, College of Health and Life Sciences, Qatar Foundation, Doha, Qatar
3 Qatar University, Department of Biomedical Science, College of Health Sciences, Doha, Qatar
4 Sidra Medicine, Division of Genetic and Genomic Medicine, Doha, Qatar
5 Sidra Medicine, Genomic Data Science Core, Doha, Qatar
6 Sidra Medicine, Department of Pediatric Nephrology, Doha, Qatar
7 Sidra Medicine, Pathology and Laboratory Medicine Department, Genetics Division, Doha, Qatar
8 Sidra Medicine, Department of Neonatology Division, Doha, Qatar
9 Sidra Medicine, Department of Pediatric Pulmonology, Doha, Qatar
10 Sidra Medicine, Department of Cardiovascular disease, Doha, Qatar
11 Sidra Medicine, Department of Pediatric Neurology, Doha, Qatar
12 Sidra Medicine, Department of Pediatric Endocrinology, Doha, Qatar
13 Sidra Medicine, Department of Pediatric Neuroradiology, Doha, Qatar
14 Weill Cornell Medicine, Department of Genetic Medicine, Qatar

Genome sequencing (GS) is a valuable diagnostic tool for patients with rare diseases, especially in populations with limited access to diagnostic services. To bridge this gap, we present initial GS findings from 250 families enrolled in a nationwide Mendelian program in Qatar, whose children were affected by rare disorders.

(continue to page 81)

(continued from page 80)

The comprehensive GS approach enabled the detection of different classes of pathogenic variants with the largest contribution from single nucleotide variants (SNVs) and small indels, 86 variants (34.4%), followed by 11 structural variations (4.4%). Additional 8 diagnoses (3.2%) were achieved through the reclassification of variants of unknown significance through co-segregation evidence, long-read sequencing, and orthogonal validation methods, contributing to an overall diagnostic yield of 42%. Moreover, 61.2% of this yield was due to recessive variants, mirroring findings from regional studies and reflecting the consanguineous nature of the population.

We identified nine novel recessive candidate genes affecting 15 individuals. These genes lacked OMIM phenotype listings and/or had only provisional gene-disease associations based on emerging evidence from the literature or collaborative platforms. Interestingly, multiple probands had dual molecular diagnoses (2.4%), mainly recessive genetic variants. We also demonstrate that genomic results led to changes in clinical management in 62.1% of cases, particularly among patients under the age of five.

As part of the study, we leveraged large family sizes to assess the value of recruiting beyond the proband. We found that including two additional siblings led to a 50% reduction in the number of rare recessive variants requiring interpretation. Furthermore, the inclusion of a second affected sibling resulted in an additional reduction, decreasing the number of variants to be assessed by up to 60%, which argues for the utility of recruiting full families at the point of care. This approach is particularly valuable in populations with high rates of inbreeding and large family sizes.

This is the first large-scale whole genome sequencing study at the point of care in a pediatric tertiary setting, and it represents a model for the future and supports the adoption of GS as a first-tier test, highlighting the contribution of our dataset to the field of genomic medicine and positioning the Middle East as a regional hub for precision medicine.

Sociodemographic factors associated with positive genome sequencing results in infants with congenital heart disease

POSTER 506

Nathaly Sweeney[1,2], Kayla Schmittau1, Karen Mestan1, Stephen Kingsmore2

1 University of California San Diego, Rady Children's Hospital, Rady Children's Institute for Genomic Medicine, San Diego, CA
2 Rady Children's Hospital, Rady Children's Institute for Genomic Medicine, San Diego, CA

Introduction: Certain sociodemographic (SD) factors are associated with a greater risk of congenital heart disease (CHD) in infants and have also been linked to poor outcomes and increased mortality in this demographic. Knowledge of SD risk factors for genetic disease can facilitate early identification of infants with CHD in need of genetic testing and potentially improve outcomes.

Methods: Retrospective evaluation of SD factors in 243 infants with CHD who underwent genomic sequencing at Rady Children's Hospital San Diego, CA, from 2017 to 2024.

Results: The overall rate of genomic disease diagnosis in the cohort was 34%. There was a trend towards higher diagnostic rate in Hispanic/Latino patients (37%) compared to White/Non-Hispanic (NH) (33%), Black/NH (29%), Asian/Pacific-Islander (22%), and Other (10%) patients. A higher rate of genomic diagnosis was associated with the primary language of Spanish (42%) as compared to English (33%) and female (40%) as compared to male (29%) patients. The proportion of genomic diagnosis was statistically significantly higher in those from areas with a bachelor's degree rate <30% (39%) compared to areas with a bachelor's degree rate >30% (24%), (p=0.0172). We also saw a trend toward greater proportion of genomic diagnosis associated with public insurance (36%) in areas with average income <$100,000 (35%); in areas with employment rate <60% (34%), although these rates did not reach statistical significance. Mortality rate in the whole cohort was 15.2% and was highest in Black/NH patients with CHD (50%). The Black/NH mortality rate was statistically significantly higher than the rate in White/NH (9%, p=<0.0001), Hispanic/Latino (12%, p=0.0001) and Asian/Pacific Islanders (22%, p=0.0439).

Conclusion: The rate of genetic diagnosis was statistically significantly higher in those from areas with a bachelor's degree rate <30%. Additionally, mortality rate was statistically significantly higher in Black/NH infants compared to White/NH, Hispanic/Latino and Asian/Pacific Islander infants. More studies are needed to elucidate potential factors leading to the observed differential increased rate of postnatal genetic disease diagnosis associated with certain sociodemographic factors in infants with CHD. Identification of these factors may prove beneficial in determining need for genetic testing postnatally but more importantly can inform programs/interventions to eliminate these disparities.

Standardizing fiber-seq for direct multi-omic chromatin profiling at single-molecule resolution

POSTER 507

Keith E. Maier[1], James T. Anderson1, Connor Fraiser1, Allison R. Hickman1, Sabrina Hunt1, Martis W. Cowles1, Bryan J. Venters1, Michael-Christopher Keogh1

1 EpiCypher, Inc., Durham, NC

Precision health requires accurate and scalable tools to understand the regulatory genome. Current methods rely on separate assays to measure chromatin accessibility, DNA methylation, nucleosome positioning, and transcription factor occupancy, preventing direct measurement of these regulatory features from the same sample or chromosome. Further, these assays depend on short-read sequencing which fragments chromatin, limiting resolution and disrupting associations between features of the same chromatin fiber.

Fiber-seq addresses these limitations by combining long-read sequencing with single-molecule chromatin accessibility profiling. The method uses an adenine methyltransferase (Hia5) to label accessible chromatin in nuclei with N6-methyladenine (6mA), which is directly detectable by platforms such as PacBio or Oxford Nanopore. This enables simultaneous readout of chromatin accessibility and native 5-methylcytosine (5mC) DNA methylation from long reads of the same DNA molecule. The chromatin accessibility labeling is remarkably sensitive, enabling inference of both nucleosome positioning and transcription factor binding. Because Fiber-seq relies on long-reads, these data can be filtered by haplotype, providing a richer, integrated view of epigenetic gene regulation than is possible with standard approaches like ATAC-seq, bisulfite sequencing, or MNase-seq.

We are currently developing Fiber-seq into a commercial assay, with a focus on enzyme standardization, automation compatibility, and robust performance across diverse sample types. Our development efforts have demonstrated high reproducibility across biological replicates, compatibility across multiple cell types, strong concordance with orthogonal short-read methods, and sufficient sensitivity to infer haplotype-specific transcription factor binding.

Fiber-seq increases the dimensionality of regulatory information obtained from long-read sequencing with a single 10-min enzymatic incubation. It is well suited for applications in biomarker discovery, pharmacogenomics, rare disease, and ascribing functional significance to non-coding variants or variants of uncertain significance. By enabling comprehensive direct multi-omic profiling from in a single assay, Fiber-seq offers a practical solution for advancing precision health research and clinical studies.

Swan Genomics: demonstration of a fundamentally new DNA sequencing technology

POSTER 509

Morassa Mohseni[1], Lawrence Lee1,2,3, David Hsu1, Min Huang1, James Brown1, Remy Ripardelli1, Niruk Shanmugam1, Simao Coelho1, Calum Fusco1, Chu Wai Liew1, Jonathan Berengut1,2, Slobodan Jergic1, Paulina Rudolffi-Soto1, Winton Wu1, Nava Whiteford1, Dale Yuzuki1, Michael Finney1, William Hyun1,4, Brock Siegel1

1 SWAN Genomics, Sydney, NSW, Australia
2 University of New South Wales, School of Biomedical Sciences, Sydney, NSW, Australia
3 University of New South Wales, ARC Centre of Excellence in Synthetic Biology, Sydney, NSW, Australia
4 University of California, San Francisco, Department Laboratory Medicine, San Francisco, CA

We report on the development and demonstration of our proprietary new sensor that has been applied to DNA sequencing. The sensor is capable of delivering high-throughput, Q30+ accuracy for both long-read and short-read DNA sequencing on a single platform, dramatically reducing costs compared to existing technologies.

The sensor comprises a plasmonic nanoantenna that enables 1000-fold enhancement of the fluorescent signal emitted from a single fluorescent nucleotide during incorporation from a DNA polymerase enzyme. Our innovation involves our patented method for the large-scale chemical synthesis of billions of plasmonic nanoantenna sensors, with a cost of goods below $1, designed specifically for their use in DNA sequencing. These sensors can be configured in a consumable device and when paired with a low-cost instrument, can achieve Q30+ DNA sequencing at a throughput that rivals or exceeds current industry leaders, with a 10x reduction in cost. Owing to the unique properties of our sensors, the platform enables read-length agnostic sequencing, accommodating DNA fragments suited for both short- and long-read sequencing applications, including those exceeding 10kb. Additionally, the sensor's unique design enables both genomics and transcriptomics analysis in a single workflow, reducing operational overhead.

We present initial data demonstrating the sensor's ability to sequence single DNA molecules. This establishes a solid foundation for ongoing technical advancements focused on refining instrument specifications and improving key metrics like read length, accuracy and throughout.

Swan's technology has the potential to revolutionize the future of genomics by making comprehensive solutions such as long-read whole genome sequencing, clinical diagnostics, and personalized medicine widely accessible at a fraction of the current cost. We hope that enabling population-scale long-read WGS will reveal strong disease correlations with WGS, supporting early interventions, improving clinical outcomes, and significantly reducing global healthcare costs.

The impact of leukocyte reduced red blood cell transfusions in neonates on germline genetic testing

POSTER 511

Sarah J. Kester[1], Sadie B. Carbajal2, Craig D. Tauscher3, Eapen K. Jacob3, Nicole J. Boczek1, Ann M. Moyer1, Maggie A. DiGuardo1,3

1 Mayo Clinic, Department of Laboratory Medicine and Pathology, Division of Laboratory Genetics and Genomics, Rochester, MN
2 Mayo Clinic, Department of Laboratory Medicine and Pathology, Division of Hematopathology, Rochester, MN
3 Mayo Clinic, Department of Laboratory Medicine and Pathology, Division of Transfusion Medicine,

Background/Objectives: Rapid genome testing is increasingly used as a diagnostic tool in the neonatal setting. However, neonates are frequently transfused, often needing multiple blood products during hospitalization. Due to the unique physiological characteristics of neonates, current transfusion practices recommend the use of leukocyte reduced (LR) red blood cells (RBCs). Given that neonates often receive product from a single donor that may comprise a larger proportion of their total blood volume, questions frequently arise about whether and to what degree a recent transfusion(s) may impact genetic testing results performed on neonatal peripheral blood specimens.

Purpose: To determine whether donor lymphocytes present in LR RBC (<5x106 WBC/unit) transfusions impact clinical molecular genetic testing in neonates, and if so, during what time-period after transfusion.

Methods: Clinical residual peripheral blood specimens were obtained from 10 neonatal patients before and after transfusion(s). Samples were de-identified and EDTA blood was extracted for DNA using the QIAGEN EZ1 Advanced XL DNA extraction or QIAsymphony DNA extraction and quantified. DNA was tested using the short tandem repeat (STR) analysis for chimerism detection to establish the relative quantity of donor and recipient DNA in a specimen (STR clinically validated lower limit of detection of 5%, but typically detects chimerism, or donor contamination, at levels as low as 1%). Single nucleotide polymorphism (SNP) analysis to generate predicted phenotypes (by analyzing the polymorphisms in blood group genes) for each blood group and human erythrocyte antigens interrogated by the test was performed to evaluate the impact on an additional molecular testing method.

Results: Neonatal patients (4-23 days old, with gestational ages ranging from 23 weeks-39 weeks) were selected who have undergone recent transfusion(s) (5 had a single LR RBC transfusion, and 5 had multiple LR RBC transfusions) and had clinical routine testing performed. Samples ranged from 2 hours to 6 days and 14 hours post-transfusion. Donor DNA was not detectable in any of the neonatal post-transfusion samples for either STR or SNP analysis revealing 0% of microchimerism post-transfusion.

Conclusions: Donor derived genomic DNA is unlikely to interfere with the interpretation of germline genetic testing in neonatal recipients following LR blood transfusions.

Transcriptomic-first approach for rare disease diagnostics and spliceopathy discovery

POSTER 512

T. M. Arriaga[1-3], R. Mendez3,4, V.S. Ganesh5-7, G. Lemire5,7, R.A. Ungar1-3, D.E. Bonner3,8, D.R. Matalon8, D. Novacic9, J.A. Bernstein3,8, Undiagnosed Diseases Network, A. O'Donnell-Luria5,7,†, M. T. Wheeler1,3,4,†, S. B. Montgomery1-3,†

1 Stanford University, Department of Genetics, Stanford, CA
2 Stanford University, Department of Pathology, Stanford, CA
3 Stanford Center for Undiagnosed Diseases, Stanford, CA
4 Stanford University, Department of Cardiology, Stanford, CA
5 Broad Institute of MIT and Harvard, Center for Mendelian Genomics, Program in Medicine and Population Genetics, Cambridge, MA
6 Brigham and Women's Hospital, Department of Neurology, Boston, MA
7 Boston Children's Hospital, Division of Genetics and Genomics, Boston, MA
8 Stanford University, Division of Medical Genetics, Department of Pediatrics, Stanford, CA
9 National Institutes of Health, National Human Genome Research Institute, Undiagnosed Diseases Program, Bethesda, MD

Spliceopathies are disorders caused by variants in the major or minor spliceosome, the latter of which excises the ~770 minor introns. They are difficult to diagnose due to the noncoding nature of spliceosomal snRNAs and their phenotypic variability. We present a transcriptome-wide approach to diagnose spliceopathies and uncover novel spliceopathies by examining patterns of minor intron retention (MIR).

We called splicing outliers using FRASER in whole-blood and fibroblast samples from 524 individuals with rare disease and familial controls from the Genomics Research to Elucidate the Genetics of Rare diseases (GREGoR) and Undiagnosed Diseases Network (UDN) consortia. Using the Minor Intron Database, we computed the number of MIR outliers per sample.

Only seven individuals from our 451-person whole-blood cohort exhibited an excess of MIR outliers (mean of 246.4 ± 96.4 vs cohort mean of 5.6 ± 32.5). Six of these were found to harbor biallelic, rare variants in minor spliceosome snRNAs: five in RNU4ATAC and one in RNU12. All six displayed features of their respective minor spliceopathy, allowing us to confirm six diagnoses and upgrade the classification of six variants, including four variants of uncertain significance (VUSs).

(continue to page 87)

(continued from page 86)

Genomic analysis of the seventh individual revealed rare (gnomAD AF <0.01, no homozygotes), highly conserved (PhyloP: 6.99, 7.69) compound heterozygous VUSs in the minor spliceosome snRNA RNU6ATAC, which was not previously associated with disease. We also identified only one individual with an excess of MIR outliers (85 vs cohort mean of 3.4 ± 4.5) in our 139-person fibroblast cohort. They also harbored rare (gnomAD AF <0.01, no homozygotes), highly conserved (PhyloP: 7.62, 3.98) compound heterozygous VUSs in RNU6ATAC. Both cases presented core features of minor spliceopathies while displaying non-overlapping symptoms, possibly due to the differing impacts of their variants on splicing.

By examining transcriptome-wide MIR outlier patterns in two clinically accessible tissues, we uncovered a candidate gene-disease relationship in RNU6ATAC, confirmed six minor spliceopathy diagnoses, and upgraded the classification of six variants. Dissecting transcriptome-

wide patterns of aberrant splicing thus represents a novel diagnostic approach for gene discovery and variant-to-function interpretation of spliceopathies.

Racial disparities and role of HPV in oropharyngeal cancer in two US populations

POSTER 604

Lawas Maria[1], Hseih M2, Patil S3, Elham Mohebbi3, Mabvakure B3, Dutteraji A3, Martinez-Cruz L 3, Garb B1, Li S1, D'Silva NJ4, Sartor1, Laura S. Rozek3

1 University of Michigan Medical School Department of Computational Medicine and Bioinformatics, Ann Arbor, MI
2 Louisiana State University Health Sciences Center School of Public Health, New Orleans, LA
3 Georgetown University, Department of Oncology, Washington, DC, USA
4 University of Michigan School of Dentistry, Ann Arbor, MI

We investigated the molecular basis of the demonstrated differences in survival from oropharyngeal squamous cell carcinoma (OPSCC) between non-Hispanic Whites (NHW) and African Americans (AA). HPV-driven (HPV+) OPSCC is associated with better treatment response and qualifies for de-escalated therapy to minimize long-term toxicity, with studies typically observing 80% or higher 5-year survival rates for HPV+ OPSCC compared to 40% or less in HPV– OPSCC. However, the predictive value of the most commonly used test, p16 immunohistochemistry (IHC), is still unclear in populations with differing HPV+ OPSCC prevalence. For example, the incidence of HPV+ OPSCC varies by geography in places like Europe or India, where southern regions have a comparably lower burden than northern regions. In this study, we analyzed two geographically and demographically diverse cohorts of AA (n = 185, from the Louisiana Tumor Registry and University of Michigan) and NHW (n = 419, from the University of Michigan) patients, performed RNA sequencing to determine HPV status of the tumors, and integrated clinical and demographic data. HPV+ prevalence was 39.5% in AA and 60.5% in NHW. We found that p16+ OPSCC was associated with significantly better survival in NHW patients (median OS not reached, i.e., >50% of patients alive at last follow-up) than in AA patients (median OS = 42.1 months, HR = 0.25, 95% CI: 0.18–0.33, $p < 0.0001$), even after adjusting for sex, age, and tumor stage. Among p16– OPSCC, AA individuals also had worse survival than NHW (median OS: 30.1 vs. 63.8 months). p16 was less predictive of HPV RNA status in AA than in NHW patients (Positive Predictive Value = 72.4% vs. 93.0%, $p = 0.006$). Among HPV RNA+ patients (68 NHW and 13 AA), AA had worse survival than NHW (median OS = 39.1 months vs. median OS not reached; 3-year survival rate 50.0% vs. 85.1%). Our findings support the need for precision diagnostics in HPV-driven OPSCC and recommend p16 not be used as a proxy for HPV status in lower-prevalence populations; in these populations, treatment decisions should be based on direct assessment of HPV.

RNA sequencing to help identify rare genetic disorders

POSTER 605

Christine Eng[1]

1 Baylor Genetics, Medical Affairs, Houston, TX

Variants of Uncertain Significance (VUS) present a stumbling block for healthcare providers and patients in their quest to get answers and avoid a lengthy, stressful, time-consuming, and expensive diagnostic odyssey. Instead of providing desperately needed information, they create uncertainty, hindering informed medical management and treatment decisions. RNAseq is a valuable functional assay that can enhance the diagnostic capability of whole genome sequencing (WGS) and whole exome sequencing (WES), and it is increasingly used to resolve challenges in interpreting genetic variants, particularly VUS. RNAseq analyzes gene expression and splicing events at the transcript level, providing crucial functional evidence that enhances variant classification by offering additional context on their impact on splicing. This tool can identify disruptions in splice site recognition, leading to abnormal protein production, and is useful for analyzing alternative splicing patterns, including exon skipping and partial or whole exon deletions. RNAseq has significantly enhanced laboratory diagnostic yield in the last decade, initially for hereditary cancer testing, but is now widely applicable across many test types and settings. Studies estimate RNAseq can reclassify between 5% and 36% of variants. Criteria for considering variants for RNAseq include classification as VUS or likely pathogenic and meeting splicing prediction algorithm criteria. In this session, Dr. Christine Eng, Baylor Genetics Chief Medical Officer, will walk through RNAseq testing criteria and recent clinical cases where the tool provided instrumental evidence in identifying pathogenic variants and uncovering novel biomarkers, leading to improved diagnosis and treatment. She will highlight how RNAseq clarifies the clinical significance of variants across the genome, facilitating prompt resolution of diagnostic odysseys, more precise patient and family management, and reducing the testing burden on patients and providers. In addition to discussing RNASeq as a tool to predict the functional significance of certain variants, she will also describe multi-omic approaches to increase the diagnostic yield of WES and WGS, such as de novo transcriptomics and metabolomics.

Traceable and medium throughput scRNA-seq of pooled individual genomes without preamplification towards clinical application

POSTER 607

Xinghua Pan[*1,2,9,15,16], Guanchuan Lin#1,2, Caiming Chen#1,2, Mengchang Xu#1,2, Bin Peng#3,4, Zhanying Dong#2, Xuewen Zhou1,2, Jinyu Gao2, Jie Yu5, Bei Jia6, Chen Luo6, Rui Hua6, Changtai Xiao2, Liyao Mai2, Yulong Zhang2, Yuanqiao He7,8, Yali Song9, Sadie L Marjani10, Xingguang Luo11, Weimin Zhang12, Mei Zhong6, Song Quan6, Sherman M Weissman13, Xin Hong14, Hao Zhu*3, Paul KH Tam*1

1 Precision Regenerative Medicine Research Centre, Medical Science Division, and State Key Laboratory of Quality Research in Chinese Medicine, Macau University of Science and Technology, Macao 999078, China.
2 Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, and Guangdong Provincial Key Laboratory of Single Cell Technology and Application, Southern Medical University, Guangzhou, Guangdong 510515, China.
3 Bioinformatics Section, School of Basic Medical Sciences, Southern Medical University, Guangzhou, Guangdong 510515, China.
4 Department of Laboratory Medicine, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong 510515, China.
5 Department of Pathology, Zhejiang University School of Medicine, Hangzhou, Zhejiang 310058, China.
6 Department of Obstetrics and Gynecology, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong 510515, China.
7 Center of Laboratory Animal Science, Nanchang University, Nanchang, Jiangxi 330031, China
8 Jiangxi Province Key Laboratory of Laboratory Animal, Nanchang University, Nanchang, Jiangxi 330031, China.
9 Center for Reproductive Medicine, Affiliated Dongguan Maternal & Child Healthcare Hospital, Southern Medical University, Dongguan, Guangdong 523000, China.
10 Department of Biology, Central Connecticut State University, New Britain, CT 06050, USA.
11 Department of Psychiatry, Yale University School of Medicine, New Haven, CT 06510, USA.
12 Institute of Cancer Research, Shenzhen Bay Laboratory, Shenzhen, Guangdong 518028, China.
13 Department of Genetics, Yale University School of Medicine, New Haven, CT 06520-8005, USA.
14 Department of Biochemistry, School of Medicine, Key University Laboratory of Metabolism and Health of Guangdong, Southern University of Science and Technology, Shenzhen, China.
15 Key Laboratory of Mental Health of China Ministry of Education, Southern Medical University, Guangzhou, Guangdong 510515, China
16 Key Laboratory of Infectious Diseases Research in South China (China Ministry of Education), Southern Medical University, Guangzhou, Guangdong 510515, China.

(continue to page 91)

(continued from page 90)

# These authors contributed equally to this article.

* Corresponding authors: E-mail: Xinghua Pan, panxinghua@most.edu.mo, ORCID: https://orcid.org/0000-0002-7421-8155; Hao Zhu, zhuhao@smu.edu.cn, ORCID: https://orcid.org/0000-0001-7384-3840; Paul KH Tam, pkhtam@must.ed··u.mo, ORCID: https://orcid.org/0000-0001-6231-3035.

Key Words:

Medium throuput scCNV-seq, Tn5 transposome, embryo mosaicism, prenatal diagnosis (PD), circulated tumor cells (CTC)

Single-cell copy number variation sequencing (scCNV-seq) is pivotal for genomic analysis across various health and disease contexts. However, existing methods often rely on whole-genome preamplification, specialized equipment, or lack traceability, posing barriers to clinical adoption. Multiplex scCNV-seq (msCNVS) enables direct barcoding individual cells in addition to the dual indexes added in later library construction. Cells are barcoded with Tn5 transposome in a microplate and pooled for streamlined processing, significantly enhancing procedure efficiency, data quality and potential throughput. msCNVS reliably distinguishes unique CNV patterns among 5 cell lines comprising 292 cells. Individual msCNVS profiles of 70 K562 cells exhibit high correlation (R=0.90-0.98) with the bulk. Triplicates of HeLa3 bulk cells demonstrate near-perfect correlation (R=0.99). The CNV profiles of primary amniotic fluid cell cultures are validated via G-banding karyotyping and chromosomal microarray analysis. msCNVS obtains superior coverage uniformity compared to MDA and MALBAC, approaching the level achieved by eMDA and DOP-PCR, and shows less fluctuation than PTA. Notably, msCNVS detects CNV deletions in two abnormal blastocysts, including one exhibiting CNV mosaicism, and reveals CNV variants in circulating tumor cells, cancerous pleural effusion cells, and patient-derived xenograft nuclei. Thus, msCNVS emerges as a robust, and highly efficient genomic testing approach for precious and rare cells typically encountered in reproductive and cancer clinics.

Index

	Last Name	First Name	Organization / Company	Page(s)
H	Han	Shunhua	Illumina, Inc.	57
K	Kester	Sarah	Mayo Clinic	85
	Khan	Umair	Bakar Computational Health Sciences Institute	67
	Kim-McManus	Olivia	Rady Children's Hospital	29
	Kirchner	Jakob	El Capitan Biosciences	71
	Korkiakoski	Anttoni	Genomill Health Inc	40, 53
	Kuo	Tsung-Ting	Yale School of Medicine	48
L	Laboy Cintrón	Dianne	University of California San Francisco	45
	LaRocca	Thomas	Stanford University School of Medicine	46
	Lawas	Maria	University of Michigan Medical School Department of Computational Medicine and Bioinformatics	88
	Lawrence	David	Centre for Cancer Biology	77
	Leung	Marco	Nationwide Children's Hospital	44
	Linder	Jodell E.	Vanderbilt University Medical Center	65
M	Maier	Keith	EpiCypher, Inc.	83
	Mamroth	Max	One Lambda Inc.	33
	Mohseni	Morassa	SWAN Genomics	84
N	Nakamura	Yuji	Rady Children's Hospital	37

	Last Name	First Name	Organization / Company	Page(s)
O	O'Donnell-Luria	Anne	Broad Institute of MIT and Harvard	36, 41, 76, 86
	Oh	Inez	WashU Medicine	64
	Oskotsky	Tomiko	University of California San Francisco	30, 38
P	Pan	Xinghua	Precision Regenerative Medicine Research Centre	90
R	Rao	Varsha	Claret Bioscience LLC	62
S	Sakr	Jasmine	University of California Irvine	68
	Sanford Kobayashi	Erica	Rady Children's Institute for Genomic Medicine	61, 79
	Sexton	Brittany	New England Biolabs, Inc.	27
	Shukor	Syukri	Hitachi High-Tech America	50
	Slaven	Neil	Lawrence Berkeley National Lab	78
	So	Austin	Clarica Genomics, Inc.	72
	Stubbs	Rebecca	Oxford Nanopore Technologies	26
	Sweeney	Nathaly	University of California San Diego	82
T	Tang	Xinyu	University of California San Francisco	30
	Tawa	Gregory	National Institutes of Health	43
	Turner	Dan	Enhanced Genomics	59

	Last Name	First Name	Organization / Company	Page(s)
V	Valdmanis	Paul	University of Washington School of Medicine	35
	Van de Wiel	Laurens	Stanford University	54
	Vollger	Mitchell	University of Utah	23
W	Weisburd	Ben	Broad Institute of MIT and Harvard	36, 76
	Wigby	Kristen	Rady Children's Institute for Genomic Medicine	56, 61, 79
	Wright	Meredith	Rady Children's Institute for Genomic Medicine	55
Z	Zhao	Grace	Agilent Technologies	34
	Zhao	Hongyu	Yale University	60
	Zhao	Boxun	Boston Children's Hospital	73, 74
	Zheng	Yu	Canal Biosciences	28
	Zhu	Rui	Yale University	49

AGBT2026™

AGRICULTURE

APRIL 12-15, 2026

SHERATON WILD HORSE PASS | PHOENIX, AZ

WWW.AGBT.ORG

AGBT2026™

PRECISION HEALTH

SEPTEMBER 14 – 16, 2026

LOEWS CORONADO BAY RESORT, SAN DIEGO, CA

KEEP YOUR EYES ON WWW.AGBT.ORG FOR MORE INFORMATION