AGBT 2026™
PRECISION HEALTH

Loews Coronado Bay Resort

San Diego, CA

September 14-16, 2026

Thank You to Our Gold Sponsor

We appreciate your generous support of the 2026 AGBT Precision Health Meeting and your commitment to advancing the future of precision health.

GeneDx

AGBT 2026™
PRECISION HEALTH

Loews Coronado Bay Resort

SEPTEMBER 14-16, 2026

Welcome to the 2026 AGBT Precision Health Meeting!

As we begin our second decade, we are grateful to the scientific community, scientific organizing committee, speakers, sponsors, and attendees whose vision and commitment have shaped this meeting into a forum for advancing genomic medicine. Your support has helped build a collaborative community dedicated to improving patient care through precision health.

This year marks an exciting new chapter. Under the leadership of our Scientific Organizing Committee, the program continues to evolve with an expanded focus on the discoveries, technologies, and collaborations that are accelerating the translation of genomics into clinical practice. From rare disease and genomic medicine to artificial intelligence, precision therapeutics, and real-world implementation, we hope the conversations you have here will inspire new ideas, partnerships, and lasting impact.

AGBT Precision Health brings together researchers, clinicians, healthcare leaders, technology innovators, biotechnology companies, diagnostics developers, and investors from around the world. By sharing knowledge across disciplines, we can accelerate the adoption of genomic technologies and improve outcomes for patients everywhere.

The AGBT Precision Health Meeting is hosted by The Genome Partnership, a not-for-profit organization dedicated to advancing research, education, collaboration, and commerce across genome science and technology.

On behalf of The Genome Partnership and the AGBT Precision Health Scientific Organizing Committee, thank you for joining us. We hope you have an engaging, productive, and inspiring meeting.

Together, we look forward to continuing to grow this community and advancing the discoveries that will shape the future of precision health.

Wendy Vlieks

Wendy Vlieks
President, The Genome Partnership
Host of the Advances in Genome Biology and Technology (AGBT) Meeting Series
AGBT General Meeting • AGBT Precision Health • AGBT Agriculture

Organizing Committee

We thank Meeting Co-Chairs Gail Jarvik, Stephen Montgomery, and Richard Wilson, along with the Scientific Organizing Committee and Contributors, for their time and dedication in developing the AGBT Precision Health 2026 program.

Bimal Chaudhari
Nationwide Children's Hospital

Wendy Chung
Boston Children's Hospital

Gail Jarvik
University of Washington, Seattle

Stephen Kingsmore
Cleveland Clinic

Heather Mefford
St. Jude Children's Research Hospital

Stephen Montgomery
Stanford University

Len Pennacchio
Lawrence Berkeley National Laboratory and DOE Joint Genome Institute

Katie Pollard
Gladstone Institutes

Heidi Rehm
Broad Institute and Massachusetts General Hospital

Jenny Taylor
University of Oxford

Richard Wilson
Nationwide Children's Hospital

Nabsys

WHOLE-GENOME STRUCTURAL VARIANT DETECTION TO ELECTRIFY YOUR RESEARCH

Electrify your SV analysis with OhmX™

Meet with our team during AGBT PH to learn how electronic genome mapping (EGM) powered by the OhmX Platform is revolutionizing structural variant detection and providing accessibility to all researchers.

Join us at our posters and presentation

- CRISPR/Cas9 optimization by electronic genome mapping
- CRISPR/Cas9 customization of electronic genome mapping repeat expansion assays
- CRISPR-assisted electronic genome mapping for robertsoniantranslocations and FSHD

September 15, 2:25-2:40 PM

Characterizing Repeat Expansions with Electronic Genome Mapping

Michael Gallagher, PhD, Software Applications Scientist, Nabsys

Book a meeting

General Information

Registration & Hospitality Desk Hours (Atrium)

- Monday, Sept. 14: 11 a.m.–7:30 p.m.
- Tuesday, Sept. 15: 7:30 a.m.–7:30 p.m.
- Wednesday, Sept. 16: 7:30 a.m.–5 p.m.

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.
- Badges are non-transferable.

Transportation

Loews Coronado Bay Resort is approximately 12 miles from San Diego International Airport (SAN), about a 20-minute drive depending on traffic. Uber, Lyft, and taxi service are readily available.

Social Events

Welcome Cocktail Hour
Monday, Sept. 14, 4–5 p.m. – Constellation Foyer

Welcome Reception and Dinner
Monday, Sept. 14, 7–9 p.m. – Bay Terrace

Women's Networking Event (RSVP required)
Tuesday, Sept. 15, 7:30–9 a.m. – Avalon

Poster Reception
Tuesday, Sept. 15, 4:30–6:30 p.m – Commodore Ballroom E

AGBT Dinner
Tuesday, Sept. 15, 6:30–8 p.m. – Constellation Ballroom & Foyer

Sponsor Social
Tuesday, Sept. 15, 8–11 p.m. – Constellation Foyer

Sips & Science: A Sunset Cruise with AGBT
Wednesday, Sept. 16, 6–9 p.m. – Loews Coronado Bay Resort Dock

General Information

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 **#AGBTPH26**.

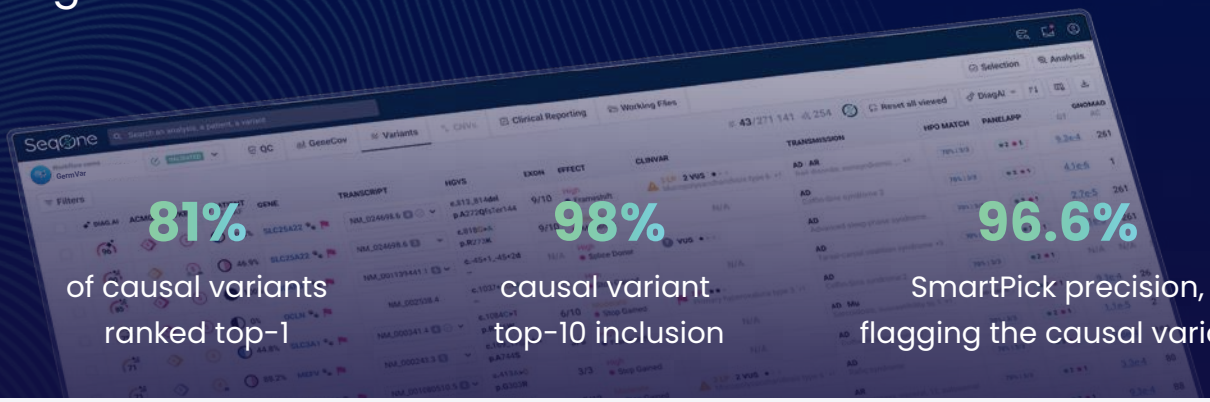
Wi-Fi Access

- Network: Illumina_Genomics
- Password: PrecisionHealth2026

SeqOne

Variant Interpretation Support, Seamlessly Automated.

Explainable, automated, and built to eliminate interpretation backlogs at scale.



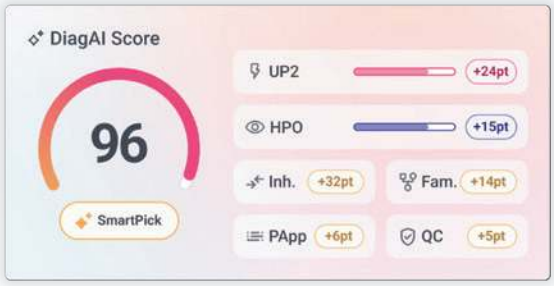
81%
of causal variants ranked top-1

98%
causal variant top-10 inclusion

96.6%
SmartPick precision, flagging the causal variant

End-to-end automation, from raw data to one actionable score.

SeqOne's **DiagAI Score** (0–100) turns complex genomic data into **one explainable, reproducible pathogenicity score.** DiagAI integrates multiple validated features, bringing automation and rigor to every analysis, **so your team can focus on the decisions only you can make.**



UP² (Universal Pathogenicity Predictor)
Instant Benchmarking: Trained on 3.6M+ ClinVar variants for robust molecular insights for SVs, CNVs and SNVs.

PhenoGenius
Automated Phenotyping: Automated phenotype-to-gene matching that collapses hours of genotype-phenoype analysis into seconds.

Expert-Driven Rules Engine
Clinical Logic at Scale: Applies ACMG-based rules automatically across every sample, fully auditable.

Genomics England 100KGP Cohort Validation
Validated on a cohort of 565 WGS singletons, across 12 disease groups.
← **Find out more in our latest white paper.**

Meet the team at Booth 202

Michael Vishnevetsky
VP, Business Development

Aaron Hersum
Account Executive

Andrew Legan
Technical FAS

SeqOne Platform is designed to support healthcare professionals in analyzing and interpreting raw genomic and sequencing data from patients undergoing NGS or CGH array analysis. By providing qualitative and actionable insights, the platform functions as a decision support system, aiding healthcare professionals in making clinical decisions. Find out more about our compliance and security framework at trust.seqone.com. Manufacturer: SeqOne (France).

Agenda

Unless otherwise noted, all sessions will take place in the Commodore Ballroom.

Monday, September 14

11:00 a.m. – 7:30 p.m.	**Hospitality Desk** Location: Atrium
Pre-Conference Programming	
1:00 p.m. – 4:00 p.m.	**Pre-Conference Workshops** Hands-on, expert-led sessions exploring tools, case studies, and practical approaches shaping precision health research and implementation.
1:00 p.m. – 2:30 p.m.	**Exposomics in Precision Health** **Speakers:** Gary Patti, Washington University in St. Louis; Chirag Patel, Harvard University; Maayan Levy, Stanford University; Donghai Liang, Emory University **Moderator:** Stephen B. Montgomery, Stanford University **Workshop Description:** This workshop explores the growing role of exposomics in advancing precision health. Speakers will present applications spanning exposure measurement, disease mapping, and emerging therapeutic approaches, followed by a panel discussion. Topics include technologies for assessing environmental exposures, progress toward a Human Exposome Project, the role of metabolites in therapy, and integrating exposure biology into precision health.

Agenda

Monday, September 14

2:30 p.m. – 4:00 p.m.

Perspectives from Expert Stakeholders Across Interventional Genomics in Rare Disease (Sponsored by Illumina)

Speakers: Petros Giannikopoulos, MD, Director, IGI Clinical Laboratory and Assistant Adjunct Professor of Laboratory Medicine, University of California, San Francisco; Olivia Kim-McManus, MD, Associate Clinical Professor, Neurosciences, University of California, San Diego; Kelley Dalby, BS, Director of Natural History and Diagnostics in Epilepsy at Praxis Precision Medicines; Lauren Black, PhD, Distinguished Scientist, Charles River.

Moderator: Katarzyna Ellsworth, PhD, FACMG, CGMB, Senior Director of Clinical Operations, Rady Children's Institute for Genomic Medicine

Workshop Description: This session covers the fast-moving field of precision genomic therapy for rare diseases, focusing on new developments in gene therapies, antisense oligonucleotides, and small molecule repurposing. It discusses how to incorporate genomic testing and data into personalized treatment plans, stressing the need for informatics, strong databases, and patient safety. Attendees will hear from geneticists, clinicians, researchers, patients, and regulators. After short presentations from each stakeholder, a fireside chat will allow for lively discussion about recent advances in creating and delivering effective treatments for people with rare conditions. Sponsored by Illumina.

Day One

Theme: Foundations and Frontiers of Genomic Medicine
This evening grounds attendees in where genomic medicine stands today and where it is headed, setting a shared framework for advances in discovery, clinical application, and emerging technologies explored throughout the meeting.

4:00 p.m. – 5:00 p.m.

Welcome Cocktail Hour
Location: Sponsor Promenade / Constellation Foyer

Session I:

Genomic Medicine Today and Tomorrow
This opening session reflects on the evolution of genomic medicine from foundational discoveries to present-day clinical impact. Speakers will examine key scientific and translational milestones while offering perspectives on the opportunities and challenges shaping the future of genomic and precision medicine.
Session Chair: Richard Wilson, Nationwide Children's Hospital (Meeting Co-Chair)

Agenda

Monday, September 14

5:00 p.m. – 5:15 p.m.	**Opening Remarks, Welcome** Richard Wilson, Nationwide Children's Hospital
	AGBT Precision Health Service Award Len Pennacchio, DOE Joint Genome Institute, Lawrence Berkeley National Laboratory **Recipient:** Michael Talkowski
5:15 p.m. – 5:45 p.m.	**Aravinda Chakravarti, NYU Grossman School of Medicine** *A scientist's perspective on Precision Medicine*
5:45 p.m. – 6:15 p.m.	**Diana Bianchi, National Institutes of Health (former)** *Prenatal Aneuploidy Screening and Incidental Detection of Cancer: Everything all at Once*
6:15 p.m. – 6:45 p.m.	**Dan Geschwind, University of California – Los Angeles** *Precision Health in a genetically diverse health system*
6:45 p.m. – 7:00 p.m.	**Christy LaFlamme, University of Washington (Abstract-Selected Talk)** *Long-read multi-ome sequencing resolves the gene regulatory mechanism of a complex structural variant causing an X-linked congenital ataxia*
7:00 p.m. – 9:00 p.m.	**Welcome Reception and Dinner** (Sponsored by SeqOne) **Location:** Bay Terrace

Tuesday, September 15

Day Two	**Theme: Intelligence, Ethics, and Discovery in Genomic Medicine** Explore how advances in artificial intelligence and computational genomics are reshaping discovery and clinical application, while also examining the ethical, regulatory, and societal considerations that accompany their adoption in precision health.
7:30 a.m. – 7:30 p.m.	**Hospitality Desk** **Location:** Atrium
7:30 a.m. – 9:00 a.m.	**Breakfast** **Location:** Commodore Terrace
7:30 a.m. – 9:00 a.m.	**AGBT Women's Networking Breakfast** (Sponsored by Illumina; RSVP Requested) Join us for the inaugural Women's Networking Event at Precision Health, a breakfast gathering focused on connection, mentorship, and shared experiences among women advancing science and precision health. **Location:** Avalon

Agenda

Tuesday, September 15

Session II:	**AI for Genomic Medicine and Diagnostics** This session examines how artificial intelligence and advanced computational methods are reshaping the interpretation of genomic data and its application in clinical care. Talks will explore AI-driven approaches to variant interpretation, gene regulation, and disease biology, with particular attention to neurological disease, as well as the translation of genomic insights into actionable diagnostics and precision medicine strategies. **Session Chair:** Katie Pollard, Gladstone Institutes
9:00 a.m. - 9:25 a.m.	**Alexis Battle, Johns Hopkins University** *Machine learning for dynamic and context-specific genetic effects*
9:25 a.m. – 9:50 a.m.	**Sara Mostafavi, Genentech** *Building AI systems for hypothesis evaluation and target assessment*
9:50 a.m. – 10:15 a.m.	**Marylyn Ritchie, Medical University of South Carolina** *Big data, real patients: AI/ML strategies for EHR-linked biobanks*
10:15 a.m. – 10:30 a.m.	**Catherine Brownstein, Boston Children's Hospital** (Abstract-Selected Talk) *Explanation-first large language model reanalysis improves diagnostic yield*
10:30 a.m. – 11 a.m.	**Coffee Break** **Location:** Sponsor Promenade / Constellation Foyer

Agenda

Tuesday, September 15

Session III:	**The Great Debate**
11:00 a.m. – 11:50 a.m.	**Modernizing Informed Consent for Genomic Medicine** As genomic medicine expands beyond specialized research settings into routine clinical care and large-scale data-sharing ecosystems, informed consent has become one of the field's most pressing, and complex, challenges. In this signature Great Debate session, panelists will examine whether current consent models are fit for purpose and explore how emerging approaches, including digital and AI-enabled tools, could support more scalable, inclusive, and trustworthy participation in precision health. The discussion will surface key tensions between efficiency and ethics, innovation and regulation, and automation and human judgment. Panelists will debate how to preserve transparency, comprehension, and patient trust while enabling broader research participation and healthcare implementation. Designed to be interactive and thought-provoking, this session invites the audience to grapple with what "modernized consent" should, and should not, look like in the next era of genomic medicine. **Moderators:** Heidi Rehm, PhD, Broad Institute of MIT and Harvard; Massachusetts General Hospital Eric Green, Chief Medical Officer, Illumina **Panelists:** Kyle Brothers, Norton Children's Research Institute Moran Snir, NEST Genomics Jenny Taylor, University of Oxford Monica Wojcik, Boston Children's Hospital
11:50 a.m. – 12:30 p.m.	**Next Gen Leadership Awards** (Sponsored by Illumina) Featuring awardees sharing their work and what the recognition means to them Gail Jarvik, University of Washington, Seattle (Meeting Co-Chair) Eric Green, Chief Medical Officer, Illumina
12:30 p.m. – 2:00 p.m.	**AGBT Lunch** Bay Terrace
1:00 p.m. – 2:00 p.m.	**Dessert with Sponsors** **Location:** Sponsor Promenade / Constellation Foyer
2:05 p.m. – 3:10 p.m.	**Sponsor Talks** Hear from industry partners advancing tools, platforms, and diagnostics supporting genomic innovation. **Moderator:** Jenny Taylor, University of Oxford
2:05 p.m. – 2:25 p.m.	**Gold Sponsor: GeneDx** Flavia Facio, Medical Affairs Director, Technologies Platform, GeneDx *From data to diagnosis: Leveraging scale, diversity, and AI-informed tools to advance variant prioritization, classification, and interpretation*

Agenda

Tuesday, September 15

2:25 p.m. – 2:40 p.m.	**Silver Sponsor: Nabsys** Michael Gallagher, PhD, Software Applications Scientist, Nabsys *Characterizing repeat expansions with electronic genome mapping*
2:40 p.m. – 2:50 p.m.	**Bronze Sponsor: SeqOne** Michael Vishnevetsky, PhD, VP of Business Development, North America, SeqOne *From platform to practice: Twelve months of US clinical deployment across the NICU, newborn screening, and community genomics.*
2:50 p.m. – 3:00 p.m.	**Bronze Sponsor: MilliporeSigma** Haiying Grunenwald, Head of Biology R&D, MilliporeSigma *After the Genome: Bringing Precision Medicine to Life with Patient-Derived Organoids*
3:00 p.m. – 3:10 p.m.	**Bronze Sponsor: Singular Genomics** Kieren Patel, Chief Business Officer, Singular Genomics *Confessions of an Instrument Maker: Thinking Outside the Box and Past the Plex*
Session IV:	**Rare Disease Mechanisms and Innovation** This session explores recent advances in understanding and addressing rare genetic diseases, highlighting progress in disease mechanism discovery, functional genomics, and the translation of insights into therapeutic strategies. Speakers will share perspectives spanning clinical investigation, experimental models, and emerging approaches to treatment. **Session Chair:** Stephen Kingsmore, Cleveland Clinic
3:10 p.m. – 3:35 p.m.	**Andrew Stergachis, University of Washington, Seattle** *You are your own best reference: Resolving the function of germline and somatic variation using donor-specific assemblies*
3:35 p.m. – 4:00 p.m.	**Monica Wojcik, Boston Children's Hospital** *New frontiers in newborn genomics: "Everything everywhere all at once"*
4:00 p.m. – 4:20 p.m.	**Poster Flash Talks** Brief Abstract-selected Poster Presentations
4:30 p.m. – 8:00 p.m.	**Poster Session, Sponsor Reception and Dinner** **Hors d'oeuvres and Posters** 4:30 p.m. – 6:30 p.m. • Odd # Posters 4:30 p.m. – 5:30 p.m. • Even # Posters 5:30 p.m. – 6:30 p.m. **Dinner and Desserts** from 6:30 p.m. – 8:00 p.m. **Location:** Commodore Ballroom E, Constellation Ballroom and Foyer
8:00 p.m. – 11:00 p.m.	**Sponsor Social** **Location:** Sponsor Promenade / Constellation Foyer

Agenda

Wednesday, September 16

Day Three	**Theme: From Insight to Impact in Rare Disease Genomic Medicine** Sessions focus on how advances in genomic medicine for rare disease move from discovery to real-world impact, spanning implementation in clinical settings, lived-patient experience, and translation into therapeutic development.
6:30 a.m. – 7:30 a.m.	**Rise & Shine with AGBT** (optional activity)
7:30 a.m. – 5:00 p.m.	**Hospitality Desk** **Location:** Atrium
7:30 a.m. – 9:00 a.m.	**Breakfast** **Location:** Commodore Terrace
Session V:	**Implementation of Precision Health and Clinical Impacts** This session examines how genomic insights are integrated into clinical care, with a focus on the roles of environment, nutrition, and early-life exposures in shaping disease risk and outcomes. Speakers will highlight emerging evidence and approaches that inform precision health strategies across the lifespan. **Session Chair:** Heather Mefford, St. Jude Children's Research Hospital
9:00 a.m. – 9:25 a.m.	**Niall Lennon, Broad Institute of Harvard and MIT** *Transformative technologies and clinical translation*
9:25 a.m. – 9:50 a.m.	**Kiran Musunuru, University of Pennsylvania** *Personalized gene-editing therapies*
9:50 a.m. – 10:15 a.m.	**Jenny Taylor, University of Oxford** *From genomic diagnostics to therapeutics: The UK's rare therapies launch pad*
10:15 a.m. – 10:40 a.m.	**Theodore G. Drivas, University of Pennsylvania** *Universal exome sequencing in critically ill adults: A diagnostic yield of 25% and race-based disparities in access to genetic testing*
10:40 a.m. – 10:55 a.m.	**Evin Padhi, Standford University** (Abstract-Selected Talk) *Biobank-scale molecular profiling refines interpretation of rare recessive alleles in Mendelian disease genes*
10:55 a.m. – 11:30 a.m.	**Coffee Break** **Location:** Sponsor Promenade / Constellation Foyer
Session VI:	**Patient's Perspective** This session brings precision health into human focus through the lived experience of a patient whose journey spans childhood cancer, survivorship, and active participation in genomic research. Joined by family and physician perspectives, the conversation explores how advances in genomic medicine shape care across time, from diagnosis and treatment to discovery and contribution, offering a moment of reflection on partnership, perseverance, and responsibility within the precision health community.

Agenda

Wednesday, September 16

11:30 a.m. – 12:15 p.m.	**Fireside Chat:** A Life Shaped by Precision Health **Moderator:** Richard Wilson, Nationwide Children's Hospital, Executive Director, The Steve and Cindy Rasmussen Institute for Genomic Medicine, Nationwide Children's Hospital **Guest(s):** Patient, her mother, and Samara Potter, Director of Clinical Cancer Genomics Outreach, Institute for Genomic Medicine, Nationwide Children's Hospital
12:15 p.m. – 1:45 p.m.	**AGBT Lunch** Bay Terrace
Session VII:	**Advances in Therapeutic Discovery to Development** This session highlights the translation of genetic insights into therapeutic strategies for rare diseases. Talks will span functional studies, target discovery, and approaches to individualized treatment, illustrating how genomic research informs the development of precision therapies. **Session Chair:** Bimal Chaudhari, Nationwide Children's Hospital
1:45 p.m. – 2:10 p.m.	**Goncalo Abecasis, Regeneron Genetics Center** *Adventures in human genetics: Improving drug discovery and development through analysis of millions of human genomes*
2:10 p.m. – 2:35 p.m.	**Alison Niewiarowska, UK Medicines and Healthcare Products Regulatory Agency (MHRA)** *Shaping the future of genomic medicine: Definitions, regulation, and equitable access*
2:35 p.m. – 3:00 p.m.	**Nadav Ahituv, University of California – San Francisco** *Functional characterization and therapeutic targeting of gene regulatory elements*
3:00 p.m. – 3:15 p.m.	**Sarah Fong, University of California – San Francisco** (Abstract-Selected Talk) *High-throughput characterization of excitatory and inhibitory gene regulatory sequences for neuronal precision gene therapy*
3:15 p.m. – 3:30 p.m.	**Xinru Zhang, Gladstone Institutes** (Abstract-Selected Talk) *Identification of early exhaustion–specific regulatory programs in T cells*
3:30 p.m. – 3:45 p.m.	**Closing Reflections** Gail Jarvik, University of Washington, Seattle (Meeting Co-Chair) Stephen Montgomery, Stanford University (Meeting Co-Chair)
6:00 p.m. – 9:00 p.m.	**Sips & Science: A Sunset Dinner Cruise with AGBTF** A crossover networking event bringing together attendees for an evening of relaxed, meaningful connection across institutions and specialties. This private yacht cruise includes dinner, drinks, and coastal sunset views, offering space to reflect on shared goals in advancing precision health and rare disease research. RSVP required; space is limited.

AGBT2027™
GENERAL MEETING

SAVE THE DATE

March 1–4, 2027

Signia by Hilton
Orlando Bonnet Creek

Orlando, FL

Learn More at www.agbt.org.

AGBT proudly recognizes Singular Genomics as a Bronze Sponsor of the 2026 AGBT Precision Health Meeting.

We are grateful for your support and partnership.

SINGULAR GENOMICS

Scientific Abstracts

POSTER

101

A multimodal precision-health framework for earlier detection of endometriosis using rare variation and unstructured electronic health record (EHR)-derived clinical phenotypes

Shefali Setia-Verma[1], Lannawill Caruth[1], Lindsay Guare[2], Jagyashila Das[1], Anaya Kapoor[3], Ananya Rajagopalan[2], Penn Medicine BioBank, Regeneron Genetics Center, Shefali Setia-Verma[1]

1. University of Pennsylvania, Department of Pathology and Laboratory Medicine, Philadelphia, PA, USA
2. University of Pennsylvania, Genomics and Computational Biology, Philadelphia, PA, USA 3University of Pennsylvania, Biomedical Informatics, Philadelphia, PA, USA

Endometriosis affects approximately 10% of reproductive-age women, yet diagnosis is frequently delayed for years, prolonging pain, impairing fertility, and increasing healthcare burden. Our recent large-scale, ancestrally diverse GWAS of endometriosis across 14 biobanks expanded the genetic landscape of disease and leveraged multi-omics to prioritize biological mechanisms, but genetic discovery alone is not sufficient for clinical risk stratification. Current prediction approaches remain limited by reliance on common variants, structured clinical data, and linear modeling, while key symptoms and diagnostic clues, including pelvic pain, dysmenorrhea, infertility, and overlapping gynecologic conditions, are often embedded in unstructured electronic health record (EHR) notes and imaging reports.

We developed a multimodal precision-health framework that integrates common variant scores, rare variant burden scores, and unstructured EHR-derived clinical phenotypes to support earlier identification of endometriosis and adenomyosis. Using the Penn Medicine BioBank, we analyzed a chart-reviewed cohort of 1,400 patients with disease status derived from unstructured EHR data and validated through detailed clinical note review. Outcomes were defined as endometriosis only, adenomyosis only, both conditions, or neither condition. We evaluated combinations of common genetic risk, rare variant burden, clinical phenotypes, and baseline covariates across random forest, XGBoost, feedforward neural network, and Gaussian Naïve Bayes models.

Multimodal models integrating EHR-derived clinical phenotypes with genetic data consistently outperformed genetics-only and covariate-only models. The strongest performance was observed for endometriosis and co-occurring endometriosis and adenomyosis, with AUROCs of approximately 0.75–0.80 across multiple feature sets and architectures. Feature sets combining clinical phenotypes with rare variant scores, with or without common variant scores, achieved the highest discrimination, whereas models relying only on rare variants or baseline covariates showed poor discrimination, with AUROCs frequently near or below 0.60.

These findings extend GWAS-based discovery toward clinically actionable risk prediction by showing that multimodal integration of rare variation and EHR-derived phenotypes can improve non-invasive risk stratification for endometriosis. Future work will incorporate multi-omics features to develop scalable screening tools that reduce diagnostic delay for an underdiagnosed and understudied gynecologic disease.

POSTER

102

A standardized multiomic liquid biopsy cryobiobanking workflow preserves cell-free DNA, extracellular vesicles, and immune cells from a single blood collection

Caden Warnick[1], Preston Shaw[1], Justin Patillo[1], David Ocampo[1], Ken Dixon[2], DianRong Tsai[1,3], Jared Barrott[1,2,3,4]

1. Brigham Young University, Cell Biology and Physiology, Provo, UT, USA
2. SpeciCare, Gainesville, GA, USA
3. Brigham Young University, Simmons Center for Cancer Research, Provo, UT, USA
4. Brigham Young University, School of Medicine, Provo, UT, USA

Background: Precision oncology increasingly relies on multimodal liquid biopsy analyses integrating circulating cell-free DNA (cfDNA), extracellular vesicles (EVs), and immune cells. However, current biobanking protocols typically optimize preservation of only a single biomarker class, limiting comprehensive downstream molecular analyses. We developed and evaluated a standardized workflow designed to preserve multiple liquid biopsy analytes from a single blood collection.

Methods: Peripheral blood was collected from healthy donors into Streck Cell-Free DNA BCT tubes and processed using SepMate™ density-gradient separation to isolate plasma and peripheral blood mononuclear cells (PBMCs). Plasma and PBMC fractions were cryopreserved independently using optimized storage conditions for 37 days before analysis. cfDNA was quantified using fluorometric assays following magnetic bead isolation, EVs were characterized by nanoparticle tracking analysis, and T-cell recovery and viability were assessed by multiparameter flow cytometry.

Results: The integrated workflow successfully preserved all three biomarker classes following cryostorage. cfDNA yields remained comparable between fresh and cryopreserved plasma, with most paired samples demonstrating equivalent or increased DNA recovery after storage. Nanoparticle tracking analysis demonstrated preservation of the characteristic exosome size distribution (30–150 nm) with similar particle population profiles before and after cryopreservation. PBMC cryopreservation maintained viable T-cell populations suitable for downstream immunophenotyping by flow cytometry. Together, these findings demonstrate that a single standardized collection and cryobiobanking workflow can preserve molecular, vesicular, and cellular liquid biopsy analytes while maintaining compatibility with downstream analytical platforms.

Conclusions: Integrated cryobiobanking enables simultaneous preservation of cfDNA, extracellular vesicles, and immune cells from a single blood draw, providing a practical framework for multimodal liquid biopsy studies and future precision oncology applications. Standardizing preservation across multiple analyte classes may facilitate longitudinal biomarker discovery, retrospective multiomic analyses, and development of comprehensive liquid biopsy assays for cancer diagnosis and therapeutic monitoring.

Advancing rare disease research through integrated data- and query-enabled discovery

POSTER

103

Gregory J. Tawa[1], Himanshu Verma[1], Jason Cheung[1], Richa Madan Lomash[1], London Toney[1], Cara Purdy[1], Sharie J. Haugabook[1], Elizabeth Ottinger[1]

1. National Center for Advancing Translational Sciences, Therapeutic Development Branch, Rockville, MD, USA

Rare disease research is often hampered by fragmented data and limited tools for connecting genetic variation to disease biology. RARe-SOURCE™, developed by NCATS (NIH National Center for Advancing Translational Science) and ABCS (Advanced Biomedical Computing Sciences), addresses this challenge by unifying rare disease datasets and enabling query-based discovery. The platform provides a single interface to link genes, variants, and diseases, supporting hypothesis generation and mechanistic insight. Through four use cases, RARe-SOURCE demonstrates its utility: identifying shared genetic architecture between rare and common diseases, modeling prevalence using allele frequency data and Hardy-Weinberg principles, uncovering common biological pathways across distinct conditions, and analyzing variant-driven disruption of protein function with structural context. By integrating siloed data and enabling flexible queries, RARe-SOURCE accelerates discovery and opens new opportunities for translational research. Future development will incorporate phenotype and chemical data, expanding its reach for precision medicine.

POSTER

104

An enhanced enzymatic methyl-sequencing (EM-seq) workflow for faster library preparation

Daniel J. Evanich[1], Alexander C. Monovich[1], Vaishnavi Panchapakesa[1], Laura N. Blum[1], Robin L. Armstrong[1], Margaret R. Heider[1], Brittany S. Sexton[1], Matthew A. Campbell[1], Bradley W. Langhorst[1], V.K. Chaithanya Ponnaluri[1], Louise Williams[1]

1. New England Biolabs, Ipswich, MA, USA

DNA methylation is an epigenetic regulator of gene expression with important functions in development and disease. NEBNext® Enzymatic Methyl-seq v2 (EM-seq™ v2) is an enzyme-based Illumina library prep workflow that enables detection of 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) by next-generation sequencing. Here we describe an updated EM-seq v2 protocol that reduces the overall EM-seq v2 reaction time to ~4.5 hours and features new recommendations for FFPE DNA inputs.

In EM-Seq v2, unmethylated cytosines are differentiated from enzymatically protected 5mC/5hmC bases through their deamination to uracil by APOBEC3A. Time savings in the updated workflow are achieved by reducing the deamination reaction time from 3 hours to 30 minutes. The modified deamination reaction is compatible with a DNA input range of 100 ng to 0.1 ng. DNA is fragmented using Covaris or NEBNext UltraShear®, a methylation compatible enzymatic fragmentation system. Illumina sequencing data shows that insert size, GC content, methylation and feature coverage are maintained between the standard (3 hour) and fast (30 min) deamination conditions.

DNA isolated from archival FFPE material is an invaluable resource for genome-wide methylation studies, however, FFPE DNA poses many notable challenges for preparing NGS libraries, such as highly variable damage. To improve DNA methylation profiling from FFPE samples, we coupled EM-Seq v2 with UltraShear fragmentation and compared to libraries generated from mechanically sheared FFPE DNA (Covaris), UltraShear fragmented EM-Seq v2 libraries featured lower duplication rates and chimeras, higher unique coverage depth and percentages of mapped reads, better uniformity, and increased library complexity.

EM-seq v2 continues to serve as a robust and cost-effective library preparation solution for accurate methylation detection. The updated EM-seq v2 reaction time of ~4.5 hours is comparable to that of TAPS+ and 5-Base workflows and results in higher throughput and faster library generation. Moreover, combining EM-seq v2 with UltraShear enzymatic fragmentation promotes further streamlining of library preparation by enabling end-to-end automation capability and greatly improves the quality of methylation profiles from FFPE DNA samples.

POSTER

105

Assessment of a large-language model (LLM) tool for analysis of genomic medicine care and outcomes within the eMERGEseq Vanderbilt cohort

Jodell E. Linder[1], Rachael Miller[1], Harris T. Bland[1], David Crosslin[2], Margaret Harr[3], Atlas Khan[4], Iftikhar Kullo[5], Nita Limdi[6], Yuan Luo[7], Brandon Buxton[1], Elizabeth M. McNally[7], Jennifer Pacheco[8], Brian Shirts[1], Shannon Terek[3], Chunhua Weng[4], Josh F. Peterson[1], Kyle W. Davis[1]

1. Vanderbilt University Medical Center, Nashville, TN, USA
2. Tulane University, New Orleans, LA, USA
3. Children's Hospital of Philadelphia, Philadelphia, PA, USA
4. Columbia University, New York, NY, USA
5. Medical College of Wisconsin, Milwaukee, WI, USA
6. University of Alabama at Birmingham, Birmingham, AL, USA
7. Northwestern University, Chicago, IL, USA
8. University of Arizona, Tucson, AZ, USA

Improving precision medicine through genomics is a central component of the eMERGE network. The eMERGEseq study sequenced and returned actionable monogenic variants for 16,218 adults. Follow-up of health care actions after return of genomic results beyond one year has not been previously reported but is critical to understanding how genomic data influences long term care. Due to the complexity of unstructured health record data, scalable methods are needed for outcome assessment. Large-language models (LLMs) may efficiently and precisely identify healthcare utilization (e.g., surveillance imaging, procedures) and new diagnoses within electronic health records (EHRs) after return of genomic risk. We use an LLM-based chart abstraction tool (Brim, using ChatGPT 4.1 mini) to extract outcomes and compare them to manual chart review (MCR). Using a random sample from one eMERGE site, we analyzed EHRs from 20 individuals with variants in 9 genes indicating high risk for breast, prostate, and colorectal cancer, hypercholesterolemia, and cardiomyopathy. Two investigators manually reviewed the same 20 charts for dates, screening tests, and procedures. We trained the LLM tool on 20 additional charts with iterative evaluation and prompt tuning, then deployed it on the original 20 manually reviewed charts. To generate the reference standard, we compared LLM and human outputs against each chart to adjudicate and re-review all findings, creating a "truth" set. The "truth" set was used to calculate precision, recall, F1 score, and accuracy. We re-reviewed human and LLM results in the EHR to define the reference standard data set of 258 adjudicated outcomes. MCR took an average of 10.5 min/chart while the LLM tool 1.6 min/chart. Compared to the reference standard, the LLM tool demonstrated precision of 0.73, recall of 0.81, F1 score of 0.77, and overall accuracy of 0.84, while human accuracy was 0.94 (both reviewers). Lessons learned include 1) iterative LLM training was time consuming, 2) certain variables required hardcoding, and 3) iterations and specific prompts were required to reduce false positives. Future work will use the LLM tool on the full cohort, and compare individuals for differences in healthcare utilization.

POSTER

106

BeginNGS: advancing newborn genome sequencing trial with retrospective federated learning models

Corrine Blucher[1], Brandan Schultz[1], Page Goddard[1], Terri Finkel[2], Ajay Talati[2], Chester W. Brown[2], Manesha Putra[3], Austin Larson[3], Katarzyna Ellsworth[1], Moran Snir[4], Laura Hayward[4], Tom Defay[5], Stephen Kingsmore[1,6], Rebecca Reimers[1]

1. Rady Children's Institute for Genomic Medicine, San Diego, CA, USA
2. The University of Tennessee Health Science Center, College of Medicine, Memphis, TN
3. University of Colorado, Boulder, CO, USA
4. NEST Genomics, New York, NY, USA
5. Alexion Pharmaceuticals Inc., Boston, MA, USA
6. Cleveland Clinic, Cleveland, OH, USA

Introduction: The BeginNGS trial uses genome sequencing (GS)-based newborn screen (NBS) to screen for 507 childhood-onset, medically actionable disorders with prospective and retrospective arms.

Methods: The eight prospective trial enrollment sites in five states approach in multiple care contexts with digital or in-person approach. Likely pathogenic or pathogenic variants are reported. All study materials were reviewed and approved by an Institutional Review Board. We tested BeginNGS query performance across expanded gene and variant spaces with additional annotation metadata and WGS biobank sequences.

Results: Prospective enrollment includes 1645 newborns. Results: 1364 negative, 71 positive, and 6 inconclusive for a positivity rate is 4.9% (71/1441) with 1.9% non-G6PDD (27/1441). One false negative screen was reported in *CYP21A2*. Illumina TruPath technology was able to detect and phase these variants in this symptomatic infant with positive biochemical screening. Federation of datasets involving over 1.2 million individuals generalized the query to different data sets, cloud platforms, and organizations with appropriate compliance and security. Additional variant metadata including allele frequency, diplotype counts, in silico and AI-score predictions, and functional evidence, improved the BeginNGS query ability to distinguish between low and high-risk screens.

Conclusions: BeginNGS has screened 1,645 participants with screen positive rate at 4.9% overall and 1.9% for non-G6PDD disorders l. Scaling of the clinical study has been facilitated by negative results reporting automation, improved variant curation through large scale data federation, and digital enrollment and results return. Federated learning models utilizing genomic data from biobanks and diagnostic testing have optimized screen performance.

Biobank-scale multi-omics for advancing precision health

POSTER

107

Tara Dutka[1], Mine Cicek[3], Hristina Denic-Roberts[1], Kim Doheny[4], Evan Eichler, Erika Faust[1], Stacey Gabriel[2], Kiran Garimella[2], Namrata Gupta[2], Elizabeth Kiernan[2], Erica Landis[1], Niall Lennon[2], Lee Lichtenstein[2], Stephen Montgomery[7], Donna Muzny[6], Evin Padhi[7], Michael Schatz[4], Sheri Schully[1], Fritz Sedlazeck[6], Michael Talkowski[2], Mary Wons[2], All of Us Research Program Long Read Working Group, All of Us Research Program Multi-omics Working Group

1. National Institutes of Health, Office of the Director, Division of Program Coordination Planning and Strategic Initiatives, All of Us Research Program, Bethesda, MD, USA
2. The Broad Institute, Cambridge, MA, USA
3. Mayo Clinic, Rochester, MN, USA
4. Johns Hopkins School of Medicine, Center for Inherited Disease Research, Baltimore, MD, USA
5. University of Washington, Seattle, WA, USA
6. Baylor College of Medicine, Human Genome Sequencing Center, Houston, TX, USA
7. Stanford University, Stanford, CA, USA

Advances in molecular methods and data generation technologies now enable large-scale biobanks to generate scalable multi-omic data types to provide layers of mechanistic insight that cannot be gleaned from genomic and health record data alone. For example, the UK Biobank partnered with industry to generate long read whole genome sequencing, metabolomics, proteomics, and methylation data at scale. Similarly, the Estonian Biobank has metabolomics data available and is now generating long read whole genome sequencing. In June of 2026, the *All of Us* Research Program released a set of overlapping multiomics on ~9k participants. These efforts are highlighting the value of multi-omics for precision medicine. Using multi-omic data we have quantified the effects of known and novel putative rare recessive disease alleles at clinically important genes. From long-read sequencing we have reconstructed novel repeat expansions and assembled clinically relevant haplotypes at loci inaccessible to prior technologies. These efforts have also highlighted challenges in generating and utilizing biobank scale multi-omic data. The technologies are still rapidly changing and for many of them standards have yet to be set by the field for what constitutes the most useful data types.This increases the risk of noisy and incompatible data when generating at scale over multiple years. Additionally, efforts to gain precision health insights from these data sets are limited by the available analytic methods, which are not yet optimized to integrate multiple data types at scale to reflect biology. The *All of Us* Research Program recently opened access to biosamples up to the wider research community for multi-omic data generation through the publication of the X01 mechanism (PAR-27-069). This mechanism opens the potential for multiple future multi-omics efforts that will need to grapple with choosing interoperable data generation methods that will produce the most insight for the research community and building cloud-based tools to handle complex multi-modal data. Here, we explore the opportunities and challenges for building multi-omic biobank data sets, emphasizing lessons learned from the *All of US* Research Program pilot multi-omic data generation and the future possibilities with broader access to biospecimens.

POSTER

108

Blood RNA signatures enable accurate discrimination of stroke subtype and onset time at hospital admission

Rashi Verma[1,2], Andrea Pearson[1], Zulma Reyes-Benitez[1], Harriet NA Blankson[1,3], Tallis Howard[1], Srikant Rangaraju[4], Alex Hall[5], Alaina Williams[5], Nicholas Stanley[5], Samayah Boynton[5], Floyd Stern[5], Tina Toosi[5], Tiera Bates[5], Jessica Garcia[5], Nicholas Liu[5], Rhidika Zakaria[5], Caidyn Ellis[5], Gavin Hurn[5], Gloria Centeno[5], Emine Guven[1,6], Theja Yelam[5], Mohamed Mubasher[7], Nirav R. Bhatt[5,8], Roger P. Simon[1], I. King Jordan[9], Michael Frankel[5], Robert Meller[1,2]

1. Morehouse School of Medicine, Neuroscience Institute, Atlanta, GA, USA
2. Morehouse School of Medicine, Institute of Translational Genomic Medicine, Atlanta, GA, USA
3. Atlanta VA Health Care System, Decatur, GA, USA
4. Yale University School of Medicine, New Haven, CT, USA
5. Emory University School of Medicine, Grady Memorial Hospital, Marcus Stroke & Neuroscience Center, Department of Neurology, Atlanta, GA, USA
6. Duzce University, Department of Biomedical Engineering, Duzce, Turkiye
7. Morehouse School of Medicine, Department of Department of Community Health and Preventive Medicine, Atlanta, GA, USA
8. University of Pittsburgh School of Medicine, Pittsburgh, PA, USA
9. Georgia Institute of Technology, School of Biological Sciences, Atlanta, GA, USA

Background: Stroke is the fourth leading cause of death and disability in the United States. Intravenous Thrombolytic (IVT) therapy are highly effective acute ischemic stroke therapies; however, their benefit is strongly time-dependent and contingent on rapid exclusion of intracranial hemorrhage. Current CT-based diagnostic paradigms can delay triage and limit access to timely therapy. This study developed and validated multivariable machine learning models using peripheral blood transcriptomes to classify stroke subtype, estimate time from onset, and predict thrombolysis eligibility.

Methods: Whole-blood samples were collected from 314 acute stroke patients at Grady Memorial Hospital (Atlanta, GA). These samples were partitioned into independent training (n=192) and validation (n=122) cohorts. Transcriptomic data were processed using the GRCh38 reference genome and quantified via StringTie2. Differentially expressed transcripts (limma) were used as predictor to train machine-learning models (caret, hidden Markov model (HMM)) in a hierarchical framework that first classified hemorrhagic stroke, then distinguished ischemic stroke from stroke mimics. We further modeled thrombolysis eligibility (≤3.5-hour window) and stratified stroke severity using the NIH Stroke Scale (NIHSS). Model performance was assessed via cross-validation and evaluated in the independent cohort using accuracy, sensitivity, specificity, and ROC-AUC metrics.

Results: Among Caret-based classifiers, GBM achieving the highest accuracy (93%). The standalone XGBoost model achieved an accuracy of 91%, whereas the HMM-based classifier demonstrated the best overall performance, achieving 100% diagnostic accuracy. A three-transcript panel achieved perfect discrimination between hemorrhagic and non-hemorrhagic strokes (100% accuracy). Furthermore, a four-transcript panel identified ischemic stroke with 97% accuracy (100% sensitivity, 96% specificity) in the validation cohort. Additional panels successfully distinguished AIS from mimics, identified patients within the critical thrombolysis window, and revealed RNA signatures that strongly correlated with clinical severity scores (NIHSS).

Conclusion: Peripheral blood RNA signatures can accurately classify stroke-subtype, time of stroke onset, and stroke severity and severity at hospital admission, supporting rapid triage and potential improvement in thrombolytic therapy access. These models warrant validation in additional cohorts.

POSTER

109

Building a production-ready framework for routine same-day ultra-rapid genome sequencing

Katie Larkin[1], Michelle Cipicchio[1], Sophie Low[1], Diana Toledo[1], Andrea Oza[1], Mark Fleharty[1], Edyta Malolepsza[1], Tom Howd[1], Corin Bolles[1], Heidi Rehm[1], Sean Hofherr[1], Niall Lennon[1]

1. Broad Clinical Labs, Burlington, MA, USA

Rapid genetic diagnosis is impactful in neonatal and pediatric intensive care units, where treatment decisions are often made within hours. Proof-of-concept studies have demonstrated whole genome sequencing (WGS) with turnaround times of less than one day, yet many approaches rely on highly specialized workflows that are difficult to operationalize in a production environment. We aim to move beyond technical demonstration to establish a production-ready framework for routine ultra-rapid WGS with return of results within 24 hours, advancing precision health by making comprehensive genomic diagnostics actionable in time-critical care.

Sequencing by Expansion (SBX) technology on the Axelios Sequencing platform enables rapid data generation using substantially shorter runtimes than traditional approaches, creating a foundation for same-day genomic analysis. An initial research study across 10 NICU research cases demonstrated turnaround times of less than 6 hours from blood sample to a variant call format (vcf) file.

Building on this technical capability, subsequent efforts focused on designing an implementation model suitable for deployment in a production laboratory. Workflow development emphasized automation, standardization, reproducibility, and minimal hands-on time. Operational considerations, including staffing models, batching, infrastructure requirements, quality management, and scalability, were evaluated to support routine clinical deployment.

The resulting framework demonstrates the feasibility of transitioning ultra-rapid WGS from an isolated technical achievement into a sustainable clinical diagnostic model. By pairing the rapid sequencing capabilities of SBX technology with a fully operationalized production workflow, this approach supports the reliable return of genomic findings within 24 hours while maintaining high-quality genomic data generation and broad variant detection capabilities. This work establishes a practical implementation model for scalable ultra-rapid genomic testing, advancing the broader adoption of rapid genomic medicine as a component of precision health in acute care settings.

POSTER

110

Capturing dGTEx developmental-age effects on gene expression using neural network sequence-to-function models

Mingyuan Li[1], Rebecca Keener[2], and Alexis Battle[2,3,4,5,6] on behalf of the dGTEx project

1. Johns Hopkins University, Department of Biology, Baltimore, MD, USA
2. Johns Hopkins University, Department of Biomedical Engineering, Baltimore, MD, USA
3. Johns Hopkins University, Department of Computer Science, Baltimore, MD, USA
4. Johns Hopkins University, Department of Genetic Medicine, Baltimore, MD, USA
5. Johns Hopkins University, Malone Center for Engineering in Healthcare, Baltimore, MD, USA
6. Johns Hopkins University, Data Science and AI Institute, Baltimore, MD, USA

Sequence-to-function (S2F) models have recently demonstrated great promise for predicting the functional consequences of DNA sequence, but they have primarily been trained using multiomic data from cell lines and adult tissues, limiting their utility for predicting gene regulatory consequences at developmental and childhood timepoints and their applicability to pediatric and rare disease. Here, we fine-tune an S2F model on the Developmental Genotype-Tissue Expression (dGTEx) resource to condition sequence-based expression prediction across tissues and developmental stages. The first phase of the dGTEx project profiled 342 bulk RNA-seq samples and 37 whole-genome sequences from 40 pediatric donors across 25 organs. Initial dGTEx analyses revealed hundreds to thousands of developmentally differentially expressed genes (dDEGs), including genes expressed in infants (0–3years old) but silenced in adolescents (15–18years old) and adults (18+ years old, evaluated in GTEx).

To capture this missing developmental context, we adapt the frozen Borzoi model using parameter-efficient low-rank adapters (LoRA) and condition predictions on tissue and developmental stage (Infant vs Adolescent). The fine-tuned model accurately recovered the direction of age-associated expression change. Among significant dDEGs, predicted changes matched the observed direction on average 83% of cases and distinguished up- from down-regulated genes with a mean AUROC of 0.88 across six organs. The model also nominates developmental differentially expressed genes *de novo*. In all organs evaluated, genes in the top quartile of predicted developmental change were 1.2 to 1.9 fold enriched for nominally significant dDEGs, suggesting that the fine-tuned model learned a regulatory axis not represented by adult-only training data. Consistent with developmental biology, predicted adolescent expression of held-out genes more closely resembled adult GTEx expression than predicted infant expression across matched tissues (mean R^2 0.78 vs 0.75).

Together, these results show that dGTEx fine-tuning enables S2F models to recover age-dependent regulation and nominate candidate dDEGs beyond the cohort's statistical power. By bringing developmental contexts into S2F modeling, this work opens a path to regulatory inference in pediatric and rare disease settings defined by small, underpowered cohorts, where recovering signals that limited sample size would otherwise miss can have a large impact.

POSTER

111

Cell-type-directed drug repurposing enables *APOE* genotype-stratified combination therapy for Alzheimer's disease

Yaqiao Li[1,2,3,*], Carlota Pereda Serras[1,2], Jessica Blumenfeld[3,4], Min Xie[3], Yanxia Hao[3,5], Elise Deng[3], You Young Chun[3], Julia Holtzman[3], Alice An[3,5], Seo Yeon Yoon[3], Xinyu Tang[1], Antara Rao[3,6], Sarah Woldemariam[1], Alice Tang[1,7], Alex Zhang[3], Jeffrey Simms[3], Iris Lo[3], Tomiko Oskotsky[1,9], Michael J Keiser[1,2], Yadong Huang[3,4,5,6,8,*], Marina Sirota[1,2,10,*]

1. University of California, San Francisco, Bakar Computational Health Sciences Institute, San Francisco, CA, USA
2. University of California, San Francisco, Pharmaceutical Science and Pharmacogenomics Graduate Program, San Francisco, CA, USA
3. Gladstone Institutes, Gladstone Institute of Neurological Disease, San Francisco, CA, USA
4. University of California, San Francisco, Neuroscience Graduate Program, San Francisco, CA, USA
5. Gladstone Institutes, Gladstone Center for Translational Advancement, San Francisco, CA, USA
6. University of California, San Francisco, Developmental and Stem Cell Biology Graduate Program, San Francisco, CA, USA
7. University of California, San Francisco, Bioengineering Graduate Program, San Francisco, CA, USA and University of California, Berkeley, CA, USA
8. University of California, San Francisco, Departments of Neurology and Pathology, San Francisco, CA, USA
9. University of California, San Francisco, Department of Medicine, Division of Clinical Informatics and Digital Transformation, San Francisco, CA, USA
10. University of California, San Francisco, Department of Pediatrics, San Francisco, CA, USA

Alzheimer's disease (AD) is heterogeneous both across patients and across cell types within a single brain. *APOE* genotype is the strongest common genetic risk factor for late-onset AD, and its effects are cell-type-specific, reshaping neuronal, astrocytic, microglial, and oligodendrocytic programs in distinct ways. Therapies aimed at a single target in a single cellular compartment have failed repeatedly, suggesting that AD requires interventions matched to both patient genotype and the distributed cellular architecture of the disease.

We developed a cell-type-directed drug repurposing framework that derives disease signatures separately for each major brain cell type from human single-nucleus RNA sequencing, stratified by *APOE* genotype, age, and sex; scores approved compounds on their ability to reverse each cell-type signature using large-scale transcriptomic perturbation libraries; and nominates drug combinations whose members jointly cover the affected cellular compartments rather than converging on a single target. Candidates are then tested in vivo and independently corroborated in real-world clinical data.

Applied to AD in a recent Cell paper, the framework nominated letrozole, the top-ranked reverser of the neuronal signature, and irinotecan, the top-ranked reverser of the glial signature. In 5xFAD x PS19 mice, the combination outperformed either monotherapy, improving spatial memory, rescued pathological deficits, and shifting cell-type-specific transcriptomic programs toward control states. Independent validation in an electronic medical record cohort of approximately 1.4 million individuals showed reduced AD incidence following exposure to each drug (letrozole, relative risk 0.466; irinotecan, relative risk 0.195).

Extending the approach to genotype-resolved hippocampal atlases of age, sex, and *APOE* responses provides the reference maps required to tailor combinations to defined patient subtypes, as new approach to precision medicine. Reversing cell-type-specific transcriptomic dysfunction with combinations of existing drugs, anchored on *APOE* genotype and corroborated in human real-world evidence, offers a tractable route to precision therapeutics for AD and a generalizable template for other genetically stratified neurodegenerative diseases.

Challenges and opportunities in expanding the variant space of newborn screening

POSTER

112

Pagé Goddard[1], William Mowrey[2], Eric Blincow[1], Katarzyna Ellsworth[1], Annette Feigenbaum[1], Michael Gelb[3], Erwin Frise[4], Christian Hansen[1], Chad Krilow[5], Jeremy Leipzig[5], Yupu Liang[2], Danny Oh[1], Stephen Pirpinias[5], Laurie Smith[1], David Dimmock[6], Monica Giovanni[7], Meredith Wright[7], Kristen Wigby[1], Stephen F. Kingsmore[1,8], Rebecca Reimers[1], Thomas Defay[2]

1. Rady Children's Institute for Genomic Medicine, San Diego, CA, USA
2. Alexion, AstraZeneca Rare Disease, Boston, MA, USA
3. Department of Biochemistry, University of Washington, Seattle, WA, USA
4. Fabric Genomics, Inc., Oakland, CA, USA
5. TileDB, Inc., Cambridge, MA, USA
6. St. George's, University of London, London, UK
7. Inflection Medicine, Boston, MA, USA
8. Cleveland Clinic, Cleveland, OH, USA

Genome-based newborn screening (gNBS) requires very low false positive rates at scale. In 2024, we developed a framework comparing observed and expected diplotype frequency across 618,290 subjects (UKB470K, the Mexico City Prospective Study, and RCIGM), spanning 342 genes and 412 disorders, to manually curate a blocklist of non-causal variants. This reduced BeginNGS false positives 48-fold in UK Biobank, from 74% to 2.0%, while preserving 99.1% recall against diagnostic genome sequencing in critically ill RCIGM patients.

Here, we present a broader study of 1,031 genes and 1,019 disorders across over 1 million subjects, with the addition of All of Us v8 and twice as many clinical genomes. The candidate variant space was also expanded to include variants of uncertain significance and conflicting pathogenicity evidence, reaching over 639,000 variants.

We built an automated, three stage blocklist pipeline, combining genotype frequency filtering, LD pruning, and diplotype contribution thresholding, to systematically screen this much larger, more complex variant space without exhaustive manual curation. Frequency filtering alone reduced the UKB positive rate to about 19%; requiring at least one ClinVar P/LP submission (228,033 variants) improved specificity to 1.9% positive, leaving 519 variants for targeted review.

Orthogonally, electronic health data from All of Us was used to distinguish non-causal variants from those with reduced penetrance or variable expressivity. In a CFTR proof of concept, EHR based phenotype scoring flagged 43 variants, shrinking the risk diplotype pool from 7,488 to 198 participants and raising the CF-diagnosis/high-phenotype-score rate from 2.8% to 56.1%. These results demonstrate a scalable path to handling variants of uncertain significance as gNBS panels expand toward comprehensive disease coverage.

POSTER

113

Convergent common- and rare-variant genetics maps lipid pathways and candidate targets for coronary artery disease

Devansh Pandey[1], **Vagheesh M. Narasimhan**[1,2,3]

1. The University of Texas at Austin, Department of Integrative Biology, Austin, TX, USA
2. The University of Texas at Austin, Oden Institute for Computational Engineering and Sciences, Austin, TX, USA
3. The University of Texas at Austin, Department of Statistics and Data Science, Austin, TX, USA

Blood lipids cause coronary artery disease (CAD), but which lipid fractions, variants, and genes carry that risk, whether common- and rare-variant signals agree, and whether the causal architecture generalizes across ancestries remain unresolved when studied one method at a time. We integrated six human-genetic analyses in a single harmonized cohort of 402,200 UK Biobank participants and 469,835 sequenced exomes, and tested every conclusion in external data.

Multivariable Mendelian randomization isolated LDL cholesterol as the strongest independent causal exposure for CAD (odds ratio 1.30 per standard deviation), with an independent triglyceride effect and no robust target-specific effect of HDL cholesterol; LDL-C and ApoB were genetically inseparable. These effects replicated in FinnGen and CARDIoGRAMplusC4D and, critically, within African-ancestry and Hispanic/admixed-American participants of the Million Veteran Program, where LDL-C remained causal after outlier-robust correction (African-ancestry odds ratio 1.52, $P = 1.3 \times 10^{-8}$).

Bayesian colocalization concentrated shared signals at five loci (PCSK9, SORT1, APOE, LDLR, TRIB1), and cis-eQTL colocalization resolved a longstanding effector-gene question by separating regulatory genes that act through expression (SORT1, APOB) from coding-driven genes with no expression signal (PCSK9, LDLR). Exome-wide rare-variant burden recovered canonical lipid genes, with LDLR loss-of-function the strongest gene-based CAD association.

To turn this evidence into candidates, we built a discovery-only convergence score that, without being shown any drug-target labels, recovered established lipid drug-target genes better than rare-variant burden alone (precision among the top ten genes 1.00 versus 0.90; area under the precision-recall curve 0.67 versus 0.57). The score also nominated the druggable phosphodiesterase PDE3B, whose loss-of-function carriers showed a severity-dependent favorable lipid profile; we show explicitly that current data are underpowered to establish CAD protection, framing PDE3B as a candidate rather than a validated target.

Together, convergent common- and rare-variant genetics maps an LDL/ApoB- and triglyceride-centered architecture of CAD that holds across cohorts and ancestries, and yields a calibrated, honestly bounded scaffold for therapeutic-target prioritization.

CRISPR/Cas9 optimization by electronic genome mapping

POSTER

114

Michael Gallagher[1], John Thompson[1], Lindsay Schneider Opsahl[1], Wentao Huang[1]

1. Nabsys, Inc., Providence, RI, USA

Cas9 binding at targeted and non-targeted sites, and its interactions with nearby proteins and DNA-binding domains, have been studied using a range of technologies. Electronic genome mapping (EGM) detects DNA nicks in long single molecules across the genome and can be used to study how gRNA mismatches affect target and off-target binding. It can also be used to examine how a DNA-bound CRISPR/dCas9 complex affects the activity of nearby enzymes.

Cas9 binding to perfectly matched gRNA sites was compared with binding to sites containing mismatches at each position in the gRNA recognition sequence. The PAM and nearby bases in the seed region were critical for binding activity. In contrast, single mismatches toward the 5′ end of the recognition sequence often allowed substantial off-target binding. The effect of individual mismatches varied between gRNAs.

EGM was also used to determine, at base-pair resolution, how close a bound CRISPR/dCas9 complex could be to a nearby nickase recognition site before nickase activity was affected. Using gRNAs that bind multiple sites throughout the human genome, we measured nickase activity at recognition sites located at different distances from the gRNA/PAM. Sites located 4–7 bp from the gRNA/PAM showed reduced nickase activity, while sites 8 bp or more away were minimally affected.

EGM can simultaneously evaluate hundreds of target and off-target sites containing single and double mismatches. These results provide further insight into Cas9 mismatch tolerance and the physical interactions between bound CRISPR/dCas9 complexes and nearby enzymes. This approach may also be useful for studying other CRISPR systems and DNA-binding proteins.

POSTER

115

Development of a multiplexed assay to characterize missense variants in *CHD2*, a chromatin remodeler and neurodevelopmental disorder (NDD)-related gene

Nicholas Bodkin[1], **Lola S. Perelman**[1], **Jeffrey D. Calhoun**[1], **Gemma L. Carvill**[1]

1. Northwestern University, Department of Neurology, Chicago, IL, USA

Background: In genetic medicine, variants of uncertain significance (VUS) pose a major challenge for families and clinicians. These DNA alterations cannot be definitively classified as benign or pathogenic, yet they make up roughly 80% of missense variants in ClinVar. Resolving VUS is essential to improving genetic care. Multiplexed Assays of Variant Effect (MAVEs) offer a high-throughput solution by assessing thousands of variants in vitro simultaneously. However, MAVEs have traditionally been used to examine oncogenes, and few have been developed to study NDD-related genes. To expand MAVEs to broader classes of genes, we developed a MAVE for *CHD2*, a dosage-sensitive chromatin remodeler, and one of the most common NDD-related genes. We discovered that *CHD2* overexpression induces slowed growth and cell death in HEK cells and leveraged this to enable a *CHD2* MAVE where cells with dysfunctional protein become enriched in the cellular population over time.

Methods: We designed a *CHD2* overexpression construct with an *attB* recombination site compatible with landing pad cell lines, such that each cell will integrate only one copy of exogenous *CHD2*. We then conducted saturation mutagenesis on six regions of the plasmid's *CHD2* coding sequence, encompassing the entire DNA-binding domain, as well as known pathogenic and benign variant "hot-spots" of the protein. We integrated 24nt barcodes into each plasmid, which added forty internal replicates to each of the 6000 amino acid changes. We then linked plasmid barcodes to variants using a novel, short-read sequencing-based method. We transfected landing pad HEK293 cells with plasmid libraries, harvested gDNA at multiple timepoints over three weeks, and sequenced barcodes.

Results: We generated functional scores for nearly all targeted amino acid substitutions (996/1000) in the first round of analysis and found perfect concordance (8/8) with pathogenic variants annotated in ClinVar. The functional scores additionally correlate well with AlphaMissense scores ($\rho = 0.487$). Finally, we have identified multiple, previously unreported variant-insensitive regions within the DNA-binding domain of *CHD2*. We expect that our novel approach, and the tools that we have developed, will accelerate the classification of variants in the many other dosage-sensitive genes implicated in NDDs.

POSTER

117

Diagnostic value of whole-genome sequencing in unresolved primary ciliary dyskinesia

Tunç Tuncel[1], Salih Berkay Berkcan[2], Ayşe Tülay Akyüz[2], Sueda Özülker[2], Sabri Aynacı[2], Adem Özleyen[2], Mehmet Ali Kök[2], Mehmet Tevfik Kara[2], Fatma Demir[3], Ömer Çağrı Akçin[3], Gülay Güleç Ceylan[4], Ahmet Cevdet Ceylan[4], Fatma Duygu Özel Demiralp[2]

1. Health Institutes of Turkiye, Ankara, Mamak, Türkiye
2. Health Institutes of Türkiye, Türkiye Biotechnology Institute, Ankara, Mamak, Türkiye
3. Ankara Bilkent City Hospital, Department of Medical Genetics, Ankara, Türkiye
4. Ankara Yıldırım Beyazit University, Department of Medical Genetics, Ankara, Türkiye

Background: Primary ciliary dyskinesia (PCD) is a genetically heterogeneous disorder in which establishing a molecular diagnosis remains challenging despite widespread use of targeted next-generation sequencing panels. Restricted gene content, uneven coverage, and difficulties detecting or interpreting non-coding, repetitive, and structurally complex variants may leave patients with a strong clinical diagnosis without a definitive molecular explanation. We evaluated the diagnostic utility of trio whole-genome sequencing (WGS) in clinically well-characterized PCD patients who remained genetically unresolved after targeted panel testing.

Methods: Three unrelated patients with well-characterized PCD and inconclusive targeted-panel results underwent family-based WGS. Each case was analyzed using both conventional Illumina short-read WGS and the Illumina TruPath Genome workflow, a proximity-mapped read technology that retains long-range information from large DNA templates while using short-read sequencing. Clinical, radiological, ciliary functional, and molecular findings were reviewed and correlated. Candidate variants were assessed for segregation, population frequency, predicted functional consequence, and consistency with the observed phenotype.

Results: WGS provided clinically meaningful molecular findings in all three patients. In patient 1, a homozygous MCIDAS frameshift variant (NM_001190787.3:c.793_812dup; p.Ser271ArgfsTer21), previously reported as heterozygous by panel sequencing, was confirmed as homozygous, refining the molecular diagnosis. In patient 2, WGS identified a homozygous intronic DNAI1 variant (NM_012144.4:c.1311+1318C>T) predicted by in silico tools to disrupt splicing, providing a more plausible explanation than a previously reported heterozygous CCDC40 variant of uncertain significance. Functional RNA studies are planned to evaluate its effect on transcript processing. In patient 3, whose targeted panel was negative, trio WGS detected two closely spaced homozygous CFAP300 indels (chr11:102058881 CTTT>C and chr11:102058887 T>TCC), in a gene absent from the original panel. TruPath Genome produced clearer support and improved local resolution for these complex adjacent indels compared with conventional short-read WGS. All patients had phenotypes consistent with PCD, including neonatal respiratory distress, chronic wet cough, recurrent sino-otopulmonary infections, bronchiectasis, and/or laterality defects.

Conclusion: Trio WGS can resolve diagnostic ambiguities caused by incorrect zygosity assignment, deep intronic variants, incomplete panel content, and complex indels. Proximity-mapped TruPath data may provide value in difficult genomic regions by improving mapping, phasing, and interpretation.

Early identification of rare disease using deep phenotyping of electronic health records

POSTER

118

Tate Tunstall[1], Pauline Ng[1], Lindsay Meyers[1], Maria-Renee Coldagelli[1], Lu Yang[1], Naveen Pereira[2], Kate Im[1], Akash Kumar[1], Matthew Rabinowitz[1]

1. MyOme, Menlo Park, CA
2. Mayo Clinic, Rochester, MN

Patients with rare genetic disorders often wait years for a correct diagnosis, highlighting the urgent need for early and efficient triage for those at risk. Electronic health records (EHRs) are rich in useful information, yet the extent to which this data is sufficient to generate prediagnostic signals is uncertain. We trained machine learning models on billing codes and phenotypes extracted from clinical notes and applied these across eight conditions in a longitudinal cohort of roughly 3 million patient records from a large health system, seeking early identification of conditions. The number of patients identified early varied by condition, ranging from 10% to 89% at 99% specificity. Median lead times exceeded one year prior to the first diagnostic billing code for most conditions, including Fabry disease (1,408 days), hereditary angioedema (2,755 days), neurofibromatosis type 1 (1,067 days), and hereditary hemorrhagic telangiectasia (2,320 days). We show that ML models trained on EHR data can prioritize patients for clinician review and confirmatory genetic testing and that the integration of diverse phenotypic data sources provides value beyond a single approach.

POSTER

119

Elucidating the functional implications of *de novo* missense variants in *KLHL20* in developmental and epileptic encephalopathies (DEE): toward therapeutic targeting by antisense oligonucleotide (ASO) therapy

Angela Gyamfi[1,2], Dianne Laboy Cintron[1], Heather C. Mefford[1]

1. St. Jude Children's Research Hospital, Department of Genomic and Translational Neuroscience, Memphis, TN, USA
2. University of Tennessee Health Science Center, Department of Genetics, Genomics, and Informatics, Memphis, TN, USA

The Kelch-like propeller protein, KLHL20, forms part of the largest subfamily of Cullin E3 ligases important for protein degradation. Activated during stress conditions, it functions to terminate autophagy with dysregulation linked to cancer and Alzheimer's disease. We previously identified heterozygous, de novo missense variants in KLHL20 in 14 individuals with a neurodevelopmental syndrome characterized by intellectual disability, epilepsy, autism spectrum disorder, hyperactivity and subtle dysmorphic facial features. Eleven (11) of these patients carry the same recurrent variant (c.1069G>A p.Gly357Arg), while 3 of them carry unique variants (c.1214G>A p.Ser405Asn; c.1262A>G p.Gln421Arg; c.1777G>Tp.Gly593Trp). Since then, there has been an additional case in the literature with infantile spasms, developmental delay, and a de novo p.Gly546Arg variant. We have also identified 6 new cases with the p.Gly357Arg variant and 1 with p.His373Gln variant; all with seizures and developmental delays. Notably all the variants cluster within the Kelch β-propeller domain, the substrate recognition domain of KLHL20. So far, the functional consequences of these variants remain unknown, and no disease model or therapeutic strategy currently exists. We hypothesize that de novo KLHL20 missense variants identified in patients act through a dominant negative gain-of-function mechanism. Specifically, we propose that KLHL20 retains its ability to bind CUL3 but interferes with substrate binding, leading to the DEE phenotype observed in patients. To test this hypothesis, we are characterizing the functional consequences of four disease-associated variants (p.Gly357Arg, p.Ser405Asn, p.Gln421Arg, p.Gly593Trp) in HEK293T cells using FLAG-tagged overexpression constructs, western blotting, immunocytochemistry and co-immunoprecipitation to assess protein abundance, localization and novel binding partners. We are generating isogenic iPSC cell lines carrying the heterozygous recurrent c.1069G>A variant by CRISPR-HDR knock-in into BJFF.6 iPSCs and will differentiate these into cortical neurons to model the disease in a physiologically relevant human cellular context. Neuronal phenotyping will include assessment of neurite morphology, PDZ-RhoGEF substrate levels and network excitability by multi-electrode array. Finally, we will design and test allele-selective ASOs targeting the c.1069G>A variant to evaluate whether specific knockdown of the mutant allele rescues the cellular phenotype.

Engineered adipose tissue as a programmable cellular pharmacy

POSTER

120

Thien Tran[1,2], Josei Sato[1,2], Nadav Ahituv[1,2]

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

Living cell-based therapeutics provide a promising frontier in medicine, yet current platforms remain limited by poor persistence, systemic toxicity, manufacturing complexity, and limited multifunctionality. Adipocytes pose as unique cellular pharmacies due to their outstanding ability to metabolize, store reagents, and secrete molecules. They are exceptionally long-lived cells, with an estimated half-life around ten years, and preclinical studies have demonstrated enzymatic secretion for more than 240 weeks after implantation. This durability distinguishes engineered adipose tissue from conventional biologics requiring repeated dosing and from cell therapies that struggle with persistence. In addition, adipocytes can be easily obtained in the clinic via liposuction and implanted via plastic surgery, creating a practical path for autologous manufacturing and localized delivery. Their natural residence in subcutaneous depots also makes them well suited for implantation as accessible, potentially retrievable therapeutic tissues.

We recently engineered adipocytes to outcompete tumors for reagents, a technology we termed adipose manipulation transplantation (AMT). We aim to extend this strategy by using AMT to deplete additional cancer-relevant nutrients and immunosuppressive metabolites. These include broad tumor metabolic dependencies, such as amino acid and lipid availability, as well as pathways such as the tryptophan–kynurenine axis. We are also engineering adipocytes to improve immunotherapy. In addition, this cellular pharmacy principle could also apply to metabolic disorders driven by toxic molecular accumulation, including maple syrup urine disease, organic acidemias, urea cycle disorders, iron overload states, and others. In these settings, adipocytes could provide continuous scavenging or secretion from a single implanted tissue rather than repeated systemic dosing.

Combined, our work aims to develop AMT as a commonly used cellular therapeutic in the clinic for a variety of diseases. By pairing durable adipose biology with programmable engineering, AMT will provide a flexible ‘cellular pharmacy’ platform for long-term treatment of cancer, metabolic disease, and other disorders requiring sustained molecular control.

POSTER

121

Enhancing short-read genome sequencing using optical genome mapping and long-read sequencing

Lisa Salz[1], Morgan Driver[1], Robert Rigobello[1], Kayla Blankenship[1], Nichole Owen[1,2], Amir Hossein Saeidian[1,2], Wenjiao Li[1,2], Katharina V. Schulze[1,2], Xiaonan Zhao[1,2]

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

Genome sequencing (GS) has become a key diagnostic tool in the evaluation of individuals with rare diseases. GS has been shown to establish molecular diagnoses at higher rates than targeted testing while reducing the need for sequential testing approaches. Because GS can increase opportunities for more informed clinical management, GS is now recommended as a first-tier test for common clinical indications, such as intellectual disability and epilepsy, by professional and medical societies. While GS is more comprehensive than targeted approaches, over 50% of individuals with a suspected genetic condition still do not receive a definitive genetic diagnosis through GS. Emerging technologies are increasingly used by clinical diagnostic laboratories to supplement short-read GS. These include optical genome mapping (OGM), which allows for detection of complex structural variants that are often missed by sequencing-based approaches, and long-read sequencing, which more accurately captures short tandem repeat (STR) expansions – limitations of current short-read GS. In March 2026, our clinical laboratory integrated OGM and long-read sequencing for assessment of these variants into our GS workflow. Here, we demonstrate the utility of these two technologies in enhancing the diagnostic capability of GS at our clinical laboratory. OGM has enabled the identification of structural variant breakpoint locations within critical genes, such as *DMD*, that are consistent with the clinical phenotype in several cases. Long-read sequencing has been utilized to confirm and more accurately size STR expansions in genes such as *GIPC1* and *RFC1*. As these emerging technologies continue to be utilized to supplement short-read GS, large-scale studies will be necessary to further demonstrate their utility in the genetic evaluation of rare diseases and support access to these technologies.

Exome findings in patients with Down syndrome

POSTER

122

April Adams[1], Kamran Hessami[1], Amilah Azam[1], Elizabeth Mizerik[1]

1. Baylor College of Medicine, Houston, TX, USA

Background: Down syndrome (DS), or trisomy 21, is associated with substantial phenotypic variability that may reflect both trisomy 21–related gene dosage effects and additional genetic variation. As genomic sequencing becomes increasingly integrated into precision medicine, identifying secondary molecular diagnoses in individuals with DS may refine prognosis, counseling, and management. We assessed the frequency and spectrum of additional genetic diagnoses among individuals with DS who underwent exome sequencing.

Methods: Electronic medical records from 2012–2026 were queried to identify patients with a confirmed diagnosis of DS who underwent exome sequencing. Abstracted data included age at DS diagnosis, age and indication for sequencing, and sequencing results. Descriptive statistics were used to summarize diagnostic yield and clinical indications.

Results: Among 4,118 patients with confirmed DS, 18 underwent exome sequencing. Six had a positive exome result, yielding a diagnostic rate of 33.3% (6/18). Diagnoses included Tourette syndrome, ZTTK syndrome, familial focal epilepsy with variable foci, NBEA-related syndrome, UFSP2-related disorder with GATA1-related disorder, and Dravet syndrome with G6PD deficiency. Most patients were diagnosed with DS in the neonatal period (10/18), with time from DS diagnosis to sequencing ranging from the same year to 14 years. Four patients were diagnosed prenatally, however four who underwent neonatal testing had prenatal suspicion for DS based on positive aneuploidy screening. Many cases had multiple indications for testing; however, the most common indications were seizure or neurologic concerns (55%), followed by immune conditions, (27%).

Conclusion: Exome sequencing identified clinically relevant secondary diagnoses in one-third of sequenced individuals with DS, particularly among those with neurologic, immune, inflammatory, or hematologic features. These findings support a precision medicine approach in which atypical or complex phenotypes in DS prompt consideration of additional genetic evaluation rather than attribution to trisomy 21 alone. Earlier genomic assessment may be especially important in the prenatal setting, where limited phenotypic information and increasing reliance on screening-based DS detection may reduce opportunities to identify coexisting genetic conditions.

FFPE preservation artifacts influence interpretation of clinically actionable variants in precision oncology

POSTER

123

Isaac Steveson[1], Ewin Jones[1], Andrew Todeschini[1], Isaac Steveson[1], Jordan Bennett[1], Ken Dixon[2], Jared Barrott[1,2,3,4]

1. Brigham Young University, Cell Biology and Physiology, Provo, UT, USA
2. SpeciCare, Gainesville, GA, USA
3. Brigham Young University, Simmons Center for Cancer Research, Provo, UT, USA
4. Brigham Young University, School of Medicine, Provo, UT, USA

Background: Clinical cancer genomics relies heavily on sequencing DNA extracted from formalin-fixed paraffin-embedded (FFPE) tissue. Although FFPE-associated sequencing artifacts are well documented, their impact on interpretation of clinically actionable variants remains incompletely characterized. We evaluated the extent to which preservation-specific sequencing discordance influences precision oncology biomarkers.

Methods: Whole-genome sequencing was performed on 50 matched cryopreserved and FFPE tumor pairs spanning 32 cancer histologies. Variant concordance between preservation methods was quantified using weighted Jaccard similarity. Preservation-specific variants were functionally annotated using Ensembl Variant Effect Predictor and evaluated with AlphaGenome tissue-specific regulatory predictions. Variants within genes associated with FDA-approved targeted therapies were interrogated using ClinVar to assess preservation-associated false-positive and false-negative pathogenic variant calls.

Results: Matched FFPE and cryopreserved specimens demonstrated an overall variant concordance of 42.2%. Preservation-specific variants differed significantly from shared variants across functional genomic annotations and predicted regulatory effects. AlphaGenome identified systematic differences in chromatin accessibility and splicing predictions for FFPE-specific variants relative to concordant variants, suggesting distinct functional characteristics of preservation-associated discordance. Analysis of clinically actionable genes identified both false-positive and false-negative pathogenic variants unique to FFPE or cryopreserved specimens, demonstrating that biospecimen preservation can alter interpretation of biomarkers used to guide targeted therapy decisions.

Conclusions: FFPE-associated sequencing artifacts extend beyond technical noise and have measurable consequences for interpretation of clinically actionable cancer variants. Combining whole-genome sequencing with regulatory effect prediction and clinical variant annotation provides a strategy for identifying preservation-associated discordance and improving confidence in precision oncology reporting. These findings support incorporation of preservation-aware computational methods into clinical genomic workflows to enhance diagnostic accuracy and therapeutic decision-making.

Founder populations in biobanks as a precision medicine engine for monogenic disease discovery

POSTER

124

Christa Caggiano[1,2], Ruhollah Shemirani[1], Priya M. Marathe[1], Jaqueline O. Odgis[1], Danielle Hoffman[1], Sarah Hanks[1], Katherine E. Bonini[1], Sabrina Suckiel[1], Michael F. Murray[1], Eimear E. Kenny[1,2,3]

1. Icahn School of Medicine at Mount Sinai, Institute for Genomic Health, New York, NY, USA
2. Icahn School of Medicine at Mount Sinai, Center for Translational Genomics, New York, NY, USA
3. Icahn School of Medicine at Mount Sinai, Department of Human Genetics, New York, NY, USA

Intro: Monogenic disease discovery often relies on specialty-care cohorts, creating ascertainment biases that limit who receives a molecular diagnosis. We hypothesized that cryptic founder structure in EHR-linked biobanks could increase power for rare-variant discovery while enabling population-based assessment of penetrance and clinical presentation.

Methods: In the Mount Sinai Million biobank (n=58,000), we identified founder groups using unsupervised identity-by-descent (IBD) clustering and developed a framework combining IBD mapping, gene-burden testing, and exome-wide association studies. We then prioritized candidate monogenic signals by integrating penetrance, phenotype rarity, and variant function, followed by clinician-guided evaluation and blinded chart review.

Results: Founder groups comprised approximately 25% of the cohort and had higher pathogenic-variant rates than non-founder groups ($P=6.6\times10^{-6}$). Across 6 founder populations and 14021 EHR-derived traits, analyses identified tens of thousands of variant-, gene-, and IBD-level associations (FDR<5%) that recovered established relationships, including *LRRK2*–Parkinson disease and *ABCB4*–liver disease, and were enriched for congenital and genetic phenotypes ($P=8.9\times10^{-5}$). The framework also revealed clinically actionable diagnostic discordance; for example, individuals homozygous for the Steel syndrome variant *COL27A1* c.2089G>C had been diagnosed with Klippel–Feil syndrome, supporting reclassification. Our prioritization framework reduced these associations to 8 founder-variant candidates for clinician-in-the-loop review. All had genotype-concordant clinical features, CADD scores >20, none were classified as pathogenic in ClinVar, and none would have been selected by P value alone. An Ashkenazi Jewish recessive variant, *LRP3* c.1846C>T, was associated with elevated LDL cholesterol (~190 mg/dL vs ~122 mg/dL in non-homozygotes) and xanthomas consistent with hyperlipidemia syndrome; a Puerto Rican recessive variant, *RIF1* c.7105C>T, tracked with increased neuropathy, muscle weakness, and ocular degeneration. Both signatures were replicated in an independent biobank (All of Us) where most phenotyped homozygotes showed the predicted phenotype (*LRP3*, 9/15 hyperlipidemia; *RIF1*, 2/3 degenerative eye disease).

Summary: Leveraging cryptic founder structure can transform diverse EHR-linked biobanks into scalable platforms for monogenic discovery, diagnostic refinement, and clinically grounded precision health.

Gene expression outlier detection of rare disease cases is limited by reference cohort heterogeneity and size

POSTER

125

Joshua L. Major-Mincer[1], Alexis C. Garretson[2], Undiagnosed Diseases Network, Aaron R. Quinlan[2]

1. University of Utah, Department of Biomedical Informatics, Salt Lake City, UT, USA
2. University of Utah, Department of Human Genetics, Salt Lake City, UT, USA

The diagnosis of rare disease cases hinges on the ability to identify variants in a patient's genome that are causal for the phenotype in question. Ascribing causation to a variant necessitates sufficient functional information regarding the effect of the variant on genes of interest. As most variants are of unknown significance, RNA-sequencing has become a popular option for functional interpretation and rare disease case resolution, since it can show whether a patient under- or over-expresses a gene relative to a reference cohort. Previous studies have included external samples from disparate sources to augment the reference cohort size when the number of available samples is insufficient, including control populations like the Genotype-Tissue Expression (GTEx) project. Including such external samples can increase noise in the cohort gene expression matrix, yet the extent to which this noise can be adequately controlled through normalization algorithms has not been extensively assessed. As a result, it is not clear how cohort heterogeneity affects outlier detection in rare disease cases.

Recognizing that sensitivity is imperative in rare disease diagnostics, we benchmarked the ability of three gene expression normalization and outlier detection models (OUTRIDER, OutSingle, and ABEILLE) to identify and control confounding noise in analysis cohorts. Specifically, we generated cohorts of variable heterogeneity, sample size, and external sample source for each Undiagnosed Diseases Network (UDN) patient that had been previously diagnosed in part by using gene expression. The recall of these positive control events was quantified as a function of these cohort characteristics. Across the tested models, we observed that outlier recall is negatively associated with heterogeneity and positively associated with cohort size. Additionally, we observed that limitations in sample size may be overcome using sample selection based on minimizing expression noise. That is, cohorts with 75 samples intentionally selected to minimize expression noise with respect to the patient had a comparable rate of recall of positive control outliers as cohorts with 200 randomly selected samples. Accordingly, we argue that confounding noise may pervade normalization procedures and thus encourage the generation of rare disease cohorts using processes informed by potential sources of confounding noise.

POSTER

126

Genome-wide characterization of FFPE sequencing artifacts reveals non-random epigenomic biases affecting somatic variant detection

Ewin Jones[1], Andrew Todeschini[1], Isaac Steveson[1], Ken Dixon[2], Jared Barrott[1,2,3,4]

1. Brigham Young University, Cell Biology and Physiology, Provo, UT, USA
2. SpeciCare, Gainesville, GA, USA
3. Brigham Young University, Simmons Center for Cancer Research, Provo, UT, USA
4. Brigham Young University, School of Medicine, Provo, UT, USA

Background: Formalin-fixed paraffin-embedded (FFPE) tissue remains the predominant biospecimen for clinical genomic sequencing despite well-recognized sequencing artifacts introduced during fixation. While artifact rates have been described, the genomic and epigenomic determinants underlying FFPE-specific variant discordance remain poorly understood. We performed a comprehensive computational analysis to characterize the biological landscape of FFPE-associated sequencing artifacts.

Methods: Whole-genome sequencing data from 50 matched cryopreserved and FFPE tumor pairs representing 32 histologic tumor types were analyzed. Variant concordance was quantified using weighted Jaccard similarity. Preservation-specific variants were integrated with Ensembl Variant Effect Predictor annotations, RepeatMasker, Roadmap Epigenomics ChromHMM states, TCGA ATAC-seq chromatin accessibility profiles, and AlphaGenome regulatory effect predictions to identify genomic features associated with preservation-specific discordance.

Results: Overall variant concordance between matched specimens was 42.2%, with substantial heterogeneity across tumors. Concordance differed significantly by mutation class, repeat element, and genomic context, indicating that preservation artifacts are non-randomly distributed across the genome. Shared variants were enriched within coding and regulatory regions, whereas FFPE-specific variants preferentially localized to repetitive genomic regions and distinct chromatin states. ChromHMM and ATAC-seq analyses demonstrated depletion of FFPE-specific variants within active regulatory elements and open chromatin. AlphaGenome predictions further identified systematic reductions in predicted chromatin accessibility for FFPE-specific variants, while splice-related functional scores differed consistently between preservation-specific and shared variant sets.

Conclusions: FFPE-induced sequencing artifacts exhibit reproducible genomic and epigenomic biases rather than random distribution. Integrating genome annotation, epigenomic context, and regulatory variant prediction provides a framework for distinguishing preservation-induced artifacts from biologically meaningful somatic variants. These findings inform development of artifact-aware analytical pipelines and improve interpretation of whole-genome sequencing data generated from routinely archived clinical specimens.

Genomic research as a critical pathway to diagnosis for individuals with rare disease

POSTER

127

Stephanie DiTroia[1], Melanie O'Leary[1], Rare Genomes Project, Anne O'Donnell-Luria[1,2], Heidi Rehm[1,3]

1. Broad Institute of MIT and Harvard, Program in Medical and Population Genetics, Cambridge, MA, USA
2. Boston Children's Hospital and Harvard Medical School, Department of Pediatrics, Division of Genetics and Genomics, Boston, MA, USA
3. Massachusetts General Hospital, Center for Genomic Medicine, Boston, MA, USA

Common barriers to a molecular diagnosis for individuals with rare disease include exhaustion of clinically available testing, limited access to specialized genetic care, and denied insurance coverage. These barriers are compounded by an incomplete understanding of the genetic basis of many rare diseases, latency between gene discovery and its adoption into diagnostic testing, and the inability to determine the functional impact of most human variation. The Rare Genomes Project (RGP) evaluates research-based genomic sequencing for undiagnosed rare disease families to address barriers to genetic testing while advancing the understanding of underlying mechanisms of rare disease. Subjects participate remotely throughout application, informed consent, sample collection, and clinical data collection. Short-read genome sequencing (GS) and family-based analysis identify known or novel causes of disease. Findings are pursued, in addition to ongoing re-analysis of unsolved families, using novel technologies and analytical approaches to maximize diagnostic yield. Returnable findings are clinically confirmed and reported via the participants' local providers. Since 2017, RGP has performed GS and analysis for 1,014 clinically heterogeneous, molecularly undiagnosed rare disease families from all fifty states, Washington DC, and Puerto Rico. Findings were returned to 279 families, including 40% with healthcare access challenges solved by known disease genes. Fifty-nine percent of returns involved one or a combination of research interventions, including the return of variants in recently discovered disease genes or the non-coding genome, functional resolution of variants of uncertain significance (VUS), and real-time candidate sharing to enable novel disease-gene discovery. Overall, this cohort achieved a returnable solve rate of 23.5% and returnable strong candidate rate of 3.5%, despite most families having had some level of prior clinical genetic testing. These results demonstrate that a research-based, patient-driven, remote genomic sequencing model can bridge the diagnostic gap for rare disease families underserved by traditional clinical care, and underscore the clinical utility of genomic sequencing and frequent re-analysis. Additionally, this study highlights the continued, critical role of research in variant interpretation, method innovation, and genomic knowledge expansion for rare disease diagnosis.

POSTER

128

High-resolution interactome constraint from 730,947 exomes maps structural signatures of purifying selection and predicts clinical variant pathogenicity

Nikolas Baya[1,2], Julia Goodrich[2], Yitang Sun[1,2], Ruchit Panchal[2], Kaitlin Samocha[1,2]

1. Massachusetts General Hospital, Center for Genomic Medicine, Boston, MA, USA
2. Broad Institute of MIT and Harvard, Cambridge, MA, USA

Gene-level constraint metrics have transformed variant interpretation, but proteins act in dynamic, multi-subunit networks. To map selective pressure across protein-protein interactions (PPIs) and complexes, we integrated PPI networks (STRING) and structural databases (CORUM, PDB) with gnomAD v4.1 constraint scores made using 730,947 exomes.

Selective pressure correlated with PPI network topology: directly interacting proteins shared more similar constraint scores than indirect interactors ($p<1e-268$ for loss-of-function and missense constraint differences), and protein eigenvector centrality correlated with constraint Z-score ($p<1e-59$). Heteromers faced stronger purifying selection than homomers ($p<3e-5$), although homomer constraint increased with subunit count ($p<0.033$), consistent with a dominant-negative-like destabilizing model. Furthermore, a derived metric quantifying protein substitution vs. conservation across CORUM complexes correlated with constraint Z-scores ($p<3e-13$), indicating structurally conserved subunits face stronger purifying selection than swappable ones.

We next estimated observed and expected variant counts in 10.7 million residues to model high-resolution constraint. Mapping inter-chain bonds from PDB for 6,101 proteins revealed that missense residues at interaction sites are more constrained than non-interface regions ($p<1e-300$). Nearly 12% (>2x expected, $p=3e-100$) of proteins showed significantly constrained missense interface residues. Interface residues were strongly enriched for ClinVar Pathogenic/Likely Pathogenic (P/LP) variants (OR=1.4, $p=4e-191$), and interface P/LP residues were more constrained than non-interface P/LP residues ($p=1.3e-4$). Notably, interface variants of uncertain significance (VUSs) were also constrained ($p=1e-70$). In a 2025-2026 retrospective ClinVar analysis, missense VUSs at interface residues were 3.4-fold more likely to be upgraded to P/LP than non-interface VUSs ($p=4.1e-4$), and 17-fold more likely if in a protein with significantly constrained missense interface residues ($p=4.8e-3$), demonstrating the predictive value of interface constraint for variant classification.

These findings establish a multi-scale structural foundation for interpreting selective constraint within protein interactions. The observed disease-relevant patterns at both protein- and residue-level resolution emphasize the power of expanding genetic context from single proteins to the broader interactome.

POSTER

129

High-throughput screens for Adipose Manipulation Transplantation (AMT) cancer therapy

Josei Sato[1,2], Kelly An[1,2], Nadav Ahituv[1,2]

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

Tumor cells promote their growth by reprogramming metabolic pathways to efficiently utilize nutrients in their surrounding microenvironment. Controlling nutrient availability within the tumor microenvironment could thus be taken advantage of for therapeutic purposes. Adipose tissue is an endocrine organ that regulates systemic energy metabolism and is routinely removed in the clinic via liposuction and implanted via plastic surgery. While white adipocytes primarily function in energy storage, brown adipocytes consume metabolic substrates to drive thermogenesis. We previously engineered white adipocytes with brown-like properties by activating *UCP1*, a central mediator of thermogenesis in brown adipocytes using CRISPR activation (CRISPRa). Furthermore, transplantation of these engineered adipocytes near tumors was shown to competitively consume metabolic substrates and suppress cancer growth, a technology we termed as Adipose Manipulation Transplantation (AMT). However, activation of *UCP1* alone primarily recapitulates the thermogenic function of endogenous brown adipocytes and might not outcompete tumors for other metabolic pathways. To identify additional metabolic candidates to outcompete tumors more rigorously, we carried high-throughput screens in adipocytes to identify candidate genes involved in glucose, fatty acid and amino acid metabolism, all three major macronutrient classes that are essential for tumor cell proliferation, whose upregulation can further suppress tumors. Combined, our approach will improve AMT as a novel cancer therapy and showcase its use for a variety of different tumors with varying metabolic parameters.

Identification of upstream stimulatory factor 1 (USF1) genetic variants associated with metabolic disease from the All of Us database

POSTER

130

Kenneth Ewool[1], Kylee Bates[1], Mary Davis[1], Julianne Grose[1]

1. Brigham Young University, Department of Microbiology and Molecular Biology, Provo, UT, USA

Precision medicine today has become an important asset in healthcare, helping tailor healthcare and treatment to individual needs. Using PheWAS, which shows how SNPs and mutant of a given gene affect a host of phenotypes, we can strengthen the use of precision health as a tool in better addressing a host of diseases such as metabolic diseases and cancer. In this study we show different SNPs of Upstream stimulatory factor 1(USF1) and how they affect lipogenesis, a key player metabolic dysregulation. USF1 is global transcription factor known to regulate several genes involved in lipid and glucose metabolism. It is particularly implicated in the disease familial combined hyperlipidemia which affects about 2% of the western population and is a major risk factor for coronary artery disease. Through the All of Us database, we have identified several SNPs of USF1 occurring in several ethnic populations which are involved in a range of disease phenotypes including cancers and metabolic diseases such as type 1 and type 2 diabetes or hyperlipidemia. Our study showed that SNPs of USF1 occurred both in the exons and introns with the exons occurring in the second and eleventh exons. Our study focused on 4 SNPs, chr1:161043331:C:T, chr1:161039890:C:T, chr1:161039365:T:A and chr1:161039677:G:A. SNP occurring at chr1:161043331:C:T occurred in the 5' UTR whilst the other 3 occurred in the 3'UTR. chr1:161043331:C:T had increased expression of USF1 both in qPCR and western blot analysis when compared to the other SNPs. This same SNP also had increased de novo lipogenesis when compared to the others. Our study shows that SNP's identified by PheWAS can display significant lipid production effects in vivo (SNP occurring at chr1:161043331:C:T), confirming the PheWAS findings and aiding in our understanding of how USF1 alleles contribute to disease whilst presenting it as a target for treating metabolic diseases.

POSTER

131

Identifying aberrant transcriptomic patterns in Undiagnosed Diseases Program (UDP) patients to enhance prioritization of candidate variants in rare disease diagnostics

Katherine Pardo[1], William Gahl[1], Barbara Swerdzewski[1], David Adams[1,2], May Malicdan[1]

1. National Institutes of Health, Undiagnosed Diseases Program, Bethesda, MD, USA
2. National Human Genome Research Institute, Office of the Clinical Director, Bethesda, MD, USA

Introduction: Transcriptomic analysis complements genomic analysis by identifying aberrant expression, aberrant splicing, mono-allelic expression, and transcriptomic structural variants implicated in rare disease. RNA sequencing (RNA-seq) outlier signals can provide more direct functional evidence than sequence-based variant effect prediction models alone. We developed an analytic framework to detect aberrant transcriptomic patterns while accounting for biological and technical variation to improve variant prioritization in rare disease evaluation.

Methods: We used version 2.6.1 of the Finding Rare Splicing Events in RNA-Seq (FRASER) tool to analyze STAR-aligned short-read RNA-Seq data from fibroblasts of patients in the Undiagnosed Diseases Program (UDP). 75 samples sequenced by the National Institutes of Health (NIH) Intramural Sequencing Center (NISC) underwent standard RNA-Seq quality control and required an RNA Integrity Number (RIN) of seven or greater. Aberrant splicing was defined as an event with a false discovery rate (FDR) < 0.1 and an absolute effect size (|delta psi|) > 0.3 for FRASER or > 0.1 for FRASER2. Six patients with Clinical Laboratory Improvement Amendments (CLIA)-validated diagnoses in OMIM disease genes associated with aberrant transcription served as positive controls, and eight unaffected biological relatives served as negative controls.

Results: All 75 samples had at least one splicing outlier, with a median of 32 aberrant splicing events per sample (mean 39.3; maximum 222), indicating substantial background burden requiring further prioritization. FRASER and FRASER2 successfully prioritized four positive-control cases involving diagnostic genes associated with global splicing dysregulation, including *DDX41, RNU4-2,* and *SOD1*. One positive control with a local splicing defect was detected however not highly prioritized. The *SNAPC4* positive control was also not strongly prioritized despite prior evidence that pathogenic *SNAPC4* variants can cause global splicing dysregulation, suggesting limited sensitivity in this cohort or tissue context. Events were prioritized based on statistical significance, effect size, read support, disease-gene association, phenotypic relevance, and overlap with rare candidate DNA variants, yielding strong diagnostic candidates.

Conclusion: Detection of aberrant transcriptomic patterns remains challenging because of technical and biological variability in RNA-seq data. Our findings support FRASER-based transcriptomic outlier analysis, integrated with genomic data, as a framework to enhance rare disease variant prioritization and potentially diagnostic yield.

POSTER

132

Identifying novel gene dysregulation associated with opioid overdose death: a meta-analysis of differential gene expression in human prefrontal cortex

Javan K. Carter[1], Bryan C. Quach[1], Caryn Willis[1], Dana B. Hancock[1], Janitza Montalvo-Ortiz[2,3], Ryan W. Logan[4,5], Consuelo Walss-Bass[6,7], Brion S. Maher[8], Eric Otto Johnson[1,9], PGC-SUD Epigenetics Working Group

1. RTI International, Omics, Epidemiology, and Analytics, Research Triangle Park, NC, USA
2. Yale School of Medicine Department of Psychiatry, Division of Human Genetics, New Haven, CT, USA
3. VA CT Healthcare System, National Center of PTSD, Clinical Neurosciences Division, West Haven and Orange, CT, USA
4. Boston University, Boston University School of Medicine, Department of Pharmacology and Experimental Therapeutics (RWL), Boston, MA, USA
5. Boston University, Center for Systems Neuroscience (RWL), Boston, MA, USA
6. University of Texas Health Science Center at Houston, McGovern Medical School, Louis A. Faillace, MD, Department of Psychiatry and Behavioral Sciences, Houston, TX, USA
7. MD Anderson Cancer Center, University of Texas Health Science Center at Houston, Graduate School of Biomedical Sciences, Houston, TX, USA
8. Johns Hopkins Bloomberg School of Public Health, Department of Mental Health, Baltimore, MD, USA
9. RTI International, Fellow Program, Research Triangle Park, NC, USA

Only recently have human postmortem brain studies of differential gene expression (DGE) associated with opioid overdose death (OOD) been published: Corradin et al. 2022; Mendez et al. 2021; Seney et al. 2021; Sosnowski et al. 2022. The four independent studies had modest sample sizes (N=40-153 decedents). We conducted a transcriptome-wide meta-analysis to increase statistical power and better understand the neurogenetics of opioid addiction. All studies had RNAseq data from human prefrontal cortex. For the meta-analysis we passed these data through consistent processing, QC, and modeling, then used a weighted Fisher's test (N=285 independent decedents; 20,098 genes) to identify OOD-associated DGE. The results identified 335 genes with significant DGE (FDR $p<0.05$). Of the 314 genes reported in prior DGE studies, 66 were retained as significant in this meta-analysis, including known genes/gene families (e.g., *OPRK1, NPAS4, DUSP, EGR*). The remaining 269 significant genes were novel (e.g., *NR4A2, SYT1, HCTRTR2*). Enrichment analyses of all three gene sets (335 complete meta-analysis genes, 66 known genes, and 269 novel genes) showed significant enrichment for the Orexin receptor pathway (Bonferroni $p=8.39e^{-4}$). Novel genes strengthened enrichment for the Orexin pathway, GO biological processes, and associated diseases, beyond the 66 previously reported genes. These associations suggest broad gene dysregulation in the prefrontal cortex and highlights the Orexin receptor pathway as a key finding in humans, providing a clear link to more than a decade of model organism studies of this pathway and addiction, including evidence of the orexin system as a promising target for opioid addiction treatment development.

POSTER

133

Identifying regulatory effects of copy number alterations in neuroblastoma using interpretable deep learning

Han Chen[1], Wendy Vlieks[2], Renata Martinez[2]

1. Salk Institute for Biological Studies, La Jolla, CA, USA
2. Advances in Genome Biology and Technology (AGBT), The Genome Partnership, St. Louis, MO, USA

Neuroblastoma (NB) is a pediatric cancer that accounts for 8% of childhood cancers and 15% of childhood cancer mortality. While NB tumors carry low single nucleotide mutation rates and few recurrent protein-coding mutations, they are characterized by frequent large-scale copy number alterations (CNAs). CNAs can influence gene expression in cis by directly altering the dosage of genes within the affected genomic region, and in trans by changing the copy number of regulatory genes whose altered expression then affects target genes elsewhere in the genome. We hypothesized that CNAs alter the expression of key regulatory genes in NB, causing transcriptome-wide expression changes that drive tumor progression. To test this, we developed a deep learning framework that predicts genome-wide RNA expression from a tumor's DNA copy-number profile, and decomposed each prediction into an interpretable cis and trans component. We trained and evaluated the model on copy number and gene expression data from 209 NB tumors from the St. Jude Children's Research Hospital and the TARGET initiative, using biologically informed gene embeddings derived from gene regulatory networks to represent conserved gene–gene relationships. We also utilized transfer learning by integrating pretrained embeddings derived from GEARS, a foundation model trained on large gene perturbation datasets that captures conserved gene-gene relationships. Incorporating these pretrained embeddings gave our model biologically informed features and significantly improved its predictive performance, achieving an average per-gene Pearson correlation of 0.49 between observed and predicted gene expression. This was a 223% improvement over our baseline model with the same architecture but non-biologically informed features (average Spearman rho of 0.22). These findings demonstrate that integrating biologically informed embeddings into our deep learning model significantly improved its ability to predict the transcriptome-wide effects of CNAs in neuroblastoma. By capturing both direct and indirect regulatory interactions, our approach allows us to identify potential key regulatory genes that drive neuroblastoma development.

IMAGINE: a model for sustainable genomic medicine: infrastructure, early outcomes, and policy considerations

POSTER

134

Ginger A. Metcalf[1], Kyler Godwin[1], Fidaa Shaib[1], Donna Muzny[1], Bo Yuan[1], Eric Venner[1], Mullai Murugan[1], Laurie Robak[1], Sandra Darilek[1], Bandana Sharma Chapagain[1], Hashem El-Serag[1], Richard Gibbs[1]

1. Baylor College of Medicine, Houston, TX, USA

As genomic sequencing becomes technically routine, the principal challenge shifts from generating genomic information to building the health system infrastructure required to deliver, steward, and continuously apply genomic information throughout the patient's lifetime. The IMAGINE Research Program at Baylor College of Medicine (BCM) is a health system scale model for this shift, embedding whole genome sequencing (WGS) screening within routine care through structured EHR integration, dual compliance consent, and a triaged return-of-results (RoR) framework.

IMAGINE operates under a single IRB approval as a hybrid research-clinical model, with one electronic consent authorizing clinical WGS testing, biospecimen and data storage for future research use, and reinterpretation of existing sequence data as new clinical indications arise. Validated results are transmitted to Epic as discrete HL7 messages rather than static reports, supporting problem-list linkage, PGx medication alerts, and future reanalysis. Participants with pathogenic/likely pathogenic (P/LP) monogenic findings (ACMG v3.2) are triaged to structured telehealth genetics consultation; pharmacogenomic (PGx) and lipoprotein(a) (LPA) results return directly via Epic.

Among 1,231 participants with completed results, 65 (5.3%) carried a P/LP variant in a medically actionable gene, 246 (20.0%) had a detected LPA risk variant, and 1,079 (87.7%) had a potentially actionable PGx phenotype. Thirty-four of the 65 ACMG-positive participants had at least 60 days of follow-up opportunity since their genetic counseling appointment; of these, 24 (71%) had completed specialty evaluation or recommended follow-up testing, an early signal of clinical uptake through the triaged RoR pathway.

This early phase of the program has highlighted policy gaps beyond program design: reimbursement models remain poorly suited to preventive genomic screening and to reanalysis over time, national genetic counseling capacity is insufficient to support population-scale RoR, and no established framework governs long term data stewardship, or payer coverage of reinterpretation. IMAGINE has developed a practical framework that integrates governance, clinical workflows, EHR infrastructure, and defined referral pathways to address many of these challenges locally. However, institution specific solutions alone will not make genomic medicine standard of care. Broad adoption will require substantial policy support, payer alignment, and financing models that recognize and sustain the full continuum of genomic care over time.

Impact of universal access to rapid genome sequencing in a level IV NICU

POSTER

135

Gwendolyn Strickland[1], Courtney P. Verscaj[1], Sophie Blumberg[1], Retna Arun[1], Rylee Kerper[1], Kai Shulman[1], Hadley Stevens Smith[1], Monica H. Wojcik[1]

1. Boston Children's Hospital, Boston, MA, USA

Background: Level IV neonatal intensive care units (NICUs) provide the highest level of neonatal care in the United States and are enriched for critically ill infants with rare diseases. Rapid genome sequencing (rGS) has high diagnostic and clinical utility in this population, but is typically offered in a selective approach per clinician discretion which may lead to delayed or missed diagnoses. We examined the impact of a universal approach to offering rGS in a level IV NICU.

Design/Methods: Eligible participants included infants without a known genetic diagnosis who were admitted to our level IV NICU for at least 48 hours. We enrolled infants into one of two cohorts over distinct 5-month periods: usual care (selective rGS) or intervention (universal rGS, all enrolled infants received rGS). We assessed parent-reported outcomes via surveys administered at enrollment and 3-month timepoints, including perceived genetic testing utility (GENE-U) and parenting stress (Parenting Stress Index).

Results: 42% (27/64) of infants in the usual care cohort received rGS and 100% (79/79) of infants in the intervention cohort received rGS. There were no observed differences in clinical or demographic features between cohorts. The diagnostic yield of rGS was similar between the intervention (21/79, 27%) and usual care (5/27, 19%) cohorts ($p = 0.45$). The intervention cohort had a higher case rate of diagnosis compared to the usual care cohort (8.4 diagnoses per 100 admissions compared to 3 diagnoses per 100 admissions, $p = 0.046$). Parent-reported stress and perceived utility of genetic testing did not differ between the cohorts at the 3-month timepoint (usual care cohort: N = 21; intervention cohort: N = 24) with additional follow-up ongoing.

Conclusions: Universal access to rGS resulted in improved diagnostic utility compared to selective access and reflected in a higher case rate of diagnosis across the level IV NICU population. Although most infants who received rGS in the intervention cohort were not suspected to have genetic conditions, the diagnostic yield of rGS was similar to the usual care cohort. Between cohorts, parents report similar perceived utility of genetic testing, and increasing access to rGS did not result in increased parental stress.

POSTER

136

Integrated epigenomic and transcriptomic signatures reveal regulatory pathways associated with cardiovascular health

Harriet N.A. Blankson[1,2], Tyler Oliver[1], Rashi Verma[1], Zulma Reyes-Benitez[1], Kimberly Rooney[2,3], Andrea Pearson[1], Peter Baltrus[4], Arshed A. Quyyumi[5,6], Priscilla Pemu[7], I. King Jordan[8], Rick Kittles[1], Herman Taylor[9], Charles D. Searles[2,3], Robert Meller[1]

1. Morehouse School of Medicine, Neuroscience Institute, Atlanta, GA, USA
2. Atlanta VA Health Care System, Decatur, GA, USA
3. Emory University School of Medicine, Division of Cardiology, Atlanta, GA, USA
4. Morehouse School of Medicine Department of Community Health and Preventive Medicine, Atlanta, GA, USA
5. Emory University, Emory Clinical Cardiovascular Research Institute, Atlanta, GA, USA
6. Emory University School of Medicine, Department of Medicine, Atlanta, GA, USA
7. Morehouse School of Medicine, Department of Medicine, Atlanta, GA, USA
8. Georgia Institute of Technology, School of Biological Sciences, Atlanta, GA, USA
9. Morehouse School of Medicine, Cardiovascular Research Institute, Atlanta, GA, USA

Background: Cardiovascular health (CVH), as defined by the American Heart Association's Life's Simple 7 (LS7) metrics, reflects the combined influence of behavioral and metabolic risk factors on cardiovascular disease. Although epigenetic mechanisms are increasingly recognized as mediators of environmental and lifestyle exposures, the regulatory epigenomic architecture underlying CVH remains incompletely characterized.

Methods: We conducted an epigenome-wide association study (EWAS) of DNA methylation using Illumina EPICv2 arrays on whole-blood samples from participants in the Morehouse-Emory Cardiovascular Center for Health Equity (MECA) cohort. Differential methylation analyses compared individuals with low-to-intermediate versus high CVH as defined by LS7 scores. Differentially methylated regions (DMRs) were identified and annotated with genomic regulatory features, including CpG island context, DNase hypersensitivity, and histone modification marks. Multi-omic integration with transcriptomic data was performed to assess concordant methylation-expression relationships. Genetic variants associated with methylation loci were further evaluated using cis-methylation quantitative trait loci (cis-meQTL) mapping and phenome-wide association studies (PheWAS).

Results: We identified 4,944 CpG sites and 7,796 DMRs significantly associated with CVH (FDR < 0.05). Hypermethylation predominated in CpG islands and promoters, whereas hypomethylation was more frequent in non-island genomic regions. Most DMRs overlapped DNase hypersensitive sites, indicating localization within regulatory chromatin. Enrichment analyses showed significant overlap with enhancer and polycomb-associated histone marks (H3K4me1 and H3K27me3) in immune and hematopoietic cell types. Integration with transcriptomic data identified genes with concordant methylation-expression relationships. These included three genes with hypomethylation-associated upregulation (*CLIC5, PTPRM,* and *SH3RF2*) and four with hypermethylation-associated downregulation (*ABCG1, ADGRA3, BTBD3,* and *PARM1*). PheWAS analysis identified 17 variants associated with clinical phenotypes, including immune and metabolic disorders.

Conclusions: These findings demonstrate that cardiovascular health is associated with coordinated epigenetic remodeling of regulatory chromatin regions linked to vascular, immune, and metabolic pathways. Integrative multi-omic analyses highlight regulatory genes that may mediate the molecular effects of lifestyle-related cardiovascular health and identify potential targets for precision prevention strategies.

Integrated proteomic analysis of molecular responses to chronic PM2.5 exposure in EGFR-mutant lung adenocarcinoma

POSTER

137

Chi-Yuan Chen[1], Chieh-Wen Chan[1], Yu-Xuan Chen[1]

1. Chang Gung University of Science and Technology, Graduate Institute of Health Industry Technology, Taoyuan, Taiwan

Background: Fine particulate matter (PM2.5) is a major environmental pollutant associated with increased lung cancer risk and poor clinical outcomes. Although PM2.5 has been implicated in therapeutic resistance, the molecular responses elicited by chronic PM2.5 exposure remain incompletely understood. This study aimed to characterize proteomic alterations induced by chronic PM2.5 exposure and identify biological pathways associated with the cellular response.

Methods: EGFR-mutant H1975 lung adenocarcinoma cells were chronically exposed to PM2.5 and evaluated for sensitivity to osimertinib. Differentially expressed proteins were identified using iTRAQ-based quantitative proteomics. Functional enrichment analyses were performed using Gene Ontology, KEGG, Reactome, and WikiPathways. Differentially expressed proteins were further mapped onto the WikiPathways Retinoblastoma (RB) signaling network. Cell-cycle distribution and DNA damage were evaluated by flow cytometry, γH2AX immunofluorescence, western blotting, and comet assays.

Results: Chronic PM2.5 exposure reduced sensitivity to osimertinib and induced extensive proteomic remodeling. Among the 112 upregulated proteins, enrichment analyses consistently identified metabolism-related pathways, including amino acid biosynthesis, serine metabolism, glycine/serine/threonine metabolism, and the Aryl hydrocarbon receptor (AhR) pathway. In contrast, the 134 downregulated proteins were significantly enriched in pathways related to cell-cycle regulation, DNA replication, DNA damage checkpoints, G1/S transition, and E2F-mediated DNA replication. Mapping the downregulated proteins onto the WikiPathways Retinoblastoma (RB) signaling network revealed alterations in key regulators of G1-S progression and DNA replication. PM2.5-treated cells also exhibited S-phase accumulation and increased DNA damage, as demonstrated by elevated γH2AX expression and increased comet tail moments.

Conclusions: Chronic PM2.5 exposure induces extensive proteomic remodeling characterized by enrichment of metabolism-related and pollutant-responsive pathways together with suppression of cell-cycle and DNA replication pathways. These molecular alterations are accompanied by S-phase accumulation, increased DNA damage, and reduced sensitivity to EGFR-TKI treatment, suggesting that coordinated metabolic adaptation and disruption of cell-cycle regulation are integral components of the cellular response to chronic PM2.5 exposure.

POSTER

138

Integrated single-cell pharmacogenomics and metabolic profiling identify precision biomarkers of measurable residual disease in multiple myeloma

Qualia Hooker[1], Sonia Ling[1,2], Cecilia Conrad[1,3], Timothy Triche Jr.[1]

1. Van Andel Research Institute, Department of Epigenetics, Grand Rapids, MI, USA
2. University of Michigan, Department of Statistics, Ann Arbor, MI, USA
3. Saint Mary's College, Department of Biology, Notre Dame, IN, USA

Measurable residual disease (MRD) following melphalan conditioning and autologous stem cell transplant (ASCT) is the strongest predictor of relapse in multiple myeloma (MM). However, the molecular determinants of persistent disease remain poorly defined. Emerging evidence suggests that metabolic reprogramming of plasma cells and immune subsets contributes to treatment resistance, while germline pharmacogenomic variation may further modulate response to melphalan and post-ASCT immune recovery. We hypothesized that integrating germline variant discovery with single-cell transcriptomics and metabolic profiling would identify precision biomarkers associated with MRD and disease progression following ASCT.

We integrated three scRNA-seq datasets comprising CD138+ cells from 54 individuals including healthy donors (n=9) , MGUS (n=14), SMM (n=19) , NDMM (n=13) alongside paired CD138- bone marrow aspirates from 28 MM patients (progressors n=13 vs. non-progressors n=15;MRD-positive n=17 vs. MRD-negative n=11) collected pre- and post-ASCT using scMerge. Germline and somatic variants were identified from single-cell transcriptomes using Monopogen and prioritized based on pharmacologically relevant loci. Cell-type specific metabolic activity was inferred from pseudobulked B-cell and malignant plasma cell populations using METAflux to identify metabolic programs that correlate with MRD and disease progression.

Integrated single-cell pharmacogenomic analysis identified 1,296 ancestry-informative variants across 18 pharmacologically relevant genes, revealing population-specific variation enriched among individuals of African ancestry (AF > 8%). Functional annotation revealed an enrichment of variants affecting energy metabolism, drug metabolism, and cell survival pathways, including loss-of-function variants in 14 genes, supporting a role for germline architecture in variability melphalan response. Cell-type metabolic flux analysis revealed distinct metabolic programs associated with treatment response. Within the B-cell compartment, antioxidant metabolism was correlated with MRD positivity (OR:1.50, 95% CI 1.05-2.16), while disease progression was associated with glycolytic flux (OR:1.44, 95% CI 1.02-2.05).

Together, these findings demonstrate that integrating germline pharmacogenomics with single-cell metabolic profiling identifies complementary transcriptomic and metabolic biomarkers that influence treatment response and highlight precision medicine strategies.

Investigating cell type–specific epigenetics of CHD2-related neurodevelopmental disorder

POSTER

139

Cyrielle Holuka[1], Christy LaFlamme[1], Edith Almanza Fuerte[1], Emily Bonkowski[1], Soham Sengupta[1], Gemma L. Carvill[2], Heather C. Mefford[1]

1. St. Jude Children's Research Hospital, Center for Pediatric Neurological Disease Research, Department of Genomic and Translational Neuroscience, Memphis, TN, USA
2. Northwestern University Feinberg School of Medicine, Department of Neurology, Chicago, IL, USA

Developmental and epileptic encephalopathies (DEEs) are a group of severe neurodevelopmental disorders characterized by early-onset, treatment-resistant seizures accompanied by developmental impairment, intellectual disability, autism spectrum disorder, and behavioral problems. Although epilepsy is a shared clinical feature of DEEs, the underlying genetic causes are highly heterogeneous, with approximately 1,500 genes implicated. Mutations in ion channels, synaptic proteins, regulators of gene expression, and regulators of neuronal excitability can lead to DEE. Pathogenic variants in *CHD2*, which regulates chromatin accessibility and gene expression, cause DEE, highlighting the role of epigenetic mechanisms in epilepsy. Epigenetic dysregulation in affected individuals with pathogenic variants results in a *CHD2*-specific DNA methylation episignature in blood-derived DNA, and this unique episignature can help clarify variant classification and diagnosis. However, the relationship of the blood-derived methylation episignature to disease pathogenesis and whether the same episignature is present in disease-relevant tissue is not known. We aim to investigate these questions using patient-derived or engineered iPSCs, neural progenitor cells (NPCs) and two-dimensional neuronal cultures. To determine whether *CHD2* patients exhibit shared or tissue-specific DNA methylation and gene expression changes, we are generating DNA methylation data including differentially methylated loci (DMLs), differentially methylated regions (DMRs), and episignatures, with RNA-seq data to assess corresponding gene expression changes. This approach enables cross-validation of epigenetic alterations with transcriptional output and allows us to determine whether the previously established *CHD2*-episignature validated in blood is conserved or cell type-specific. DNA methylation profiling of blood-derived samples from 68 individuals with variants in *CHD2* using EPIC v1 and v2 arrays identified DMLs and DMRs. Additionally, the *CHD2*-associated episignature was observed in blood-derived DNA. However, analysis of methylation data from NPC (n=12) and iPSC models (n=24) with *CHD2* variants showed distinct methylation profiles. The results suggest cell type-specific methylation patterns that require further investigation. Analysis of RNA-seq data from NPC and iPSC models will determine whether cell type-specific DMRs are associated with gene expression changes. Integration of methylation and RNAseq data across relevant cell types will advance our understanding of disease pathogenesis, enable robust biomarker discovery, and facilitate the identification of both shared and tissue-specific therapeutic targets aligned with clinical heterogeneity.

POSTER

140

Joint single-cell capture of Cas9 edits and transcriptomes reveals regulatory effects of on- and off-target editing

Graham McVicker[1,#], Mickey Lorenzini[1,2,*], Brad Balderson[1,*], Shanna N. Lavalle[1], Karthyayani Sajeev[1], Aaron J. Ho[1], Madelyn Light[3], Olalekan H. Usman[3], Gavin Kurgan[3]

1. 1Salk Institute for Biological Studies, Integrative Biology Laboratory, La Jolla, CA, USA
2. 2University of California, San Diego, Biomedical Sciences Graduate Program, La Jolla, CA, USA
3. 3Integrated DNA Technologies, Coralville, IA, USA

* Lead authors

Presenting and corresponding author

CRISPR/Cas9 has broad applications in therapeutic genome editing and functional genomics, but a major limitation is variable on-target activity and pervasive off-target editing. Off-target editing poses safety risks for CRISPR therapies because edits at unintended genomic sites can disrupt gene regulation or cause genotoxicity. Bulk off-target detection methods such as GUIDE-seq and DISCOVER-seq identify on- and off-target edit sites genome-wide but do not connect them to changes in gene expression. Perturb-seq associates perturbations with single-cell gene expression changes, but because it relies on guide RNA detection rather than direct edit detection, off-target events are invisible to it.

To address these problems, we have developed Superb-seq, a three-step method for joint single-cell capture of genome-wide Cas9 edits and transcriptomes. First, we edit live cells with Cas9 ribonucleoprotein and label edits with an optimized donor DNA that contains a T7 promoter sequence. Second, we perform T7 in situ transcription on fixed cells to generate edit-reporting transcripts. Third, we capture T7 transcripts and endogenous transcripts by combinatorial single-cell RNA-seq.

We performed Superb-seq on 35,000 cells from three cell lines. In K562 cells, we used seven guide RNAs to target four chromatin remodeler genes and detected 43 edit sites, including 36 off-target sites, across 5,761 edited cells (up to 5 edits per cell). We verified edits with amplicon sequencing (rhAmp-seq), and confirmed accurate quantification of edit frequency (Pearson's $R = 0.87$, $P = 3.7 \times 10^{-10}$). By using edit status rather than guide identity, Superb-seq improved estimation of downstream gene expression effects compared to Perturb-seq (log2 OR 0.66 vs. −0.06, $P = 0.0015$). In K562, HEK293T, and U2OS cells, we compared Superb-seq against four bulk edit detection methods using two well-characterized guide RNAs (VEGFA site 2 and HEK site 4). We detected 210 and 27 high-confidence off-target sites (≥2 orthogonal methods) compared to 152 and 134 by GUIDE-seq. We also identified 19 non-coding off-target edits associated with differential gene expression.

In summary, Superb-seq identifies on- and off-target genome edits and determines which edits affect gene expression, providing a scalable method to assess the safety of CRISPR-based therapeutics.

POSTER

141

Load, scan, purify: automated, reproducible genomic DNA extraction from different clinical sample types

Shariya Kennedy[1], **Jared Bailey**[1]

1. Omega Bio-tek, Norcross, GA, USA

With increasing processing demand in precision health and clinical laboratories, genomic DNA extraction workflows that are both scalable and reliable while minimizing variability are more important than ever. The Mag-Bind® Blood & Saliva DNA LSP Kit enables automated, high-throughput genomic DNA purification from blood, saliva, buffy coat, and cultured cells using pre-filled, barcoded reagents compatible with open liquid handling platforms, enabling standardization of genomic DNA extraction across the lab with one kit for the most common sample types. The workflow follows a "Load, Scan, Purify" approach where sealed reagent reservoirs, tubes, and plastic ware are identified and volume-verified prior to run initiation, streamlining deck setup, reducing operator variability, and decreasing setup time by up to 50%. This kit can process up to four 96-well plates (384 samples) in ~2.5 hours with minimal hands-on time. Performance was evaluated by subjecting genomic DNA extracted from saliva and whole blood to NanoDrop® analysis. This kit consistently generated high yields of high-quality DNA with A260/A280 ratios ~1.8, A260/A230 ratios ~2.0, and low variability across plates. qPCR analysis demonstrated expected amplification efficiency of ~3.3 per 10-fold dilution, confirming inhibitor-free DNA suitable for downstream applications. These results demonstrate that the Mag-Bind® Blood & Saliva DNA LSP Kit enables scalable, walkaway automation while maintaining DNA quality suitable for sensitive downstream applications.

Long-read sequencing for improved methylation classification of Mendelian conditions

POSTER

142

Sophia B. Gibson[1], **Miranda P.G. Zalusky**[2,3], **Zachary B. Anderson**[2,3], **Joy Goffena**[2,3], **Sophie HR Storz**[2,3], **Yuwei Shi**[4], **Kristina Lanko**[4], **Samantha Vergano**[2], **Stefan Barakat**[4], **Danny E. Miller**[1,2,3,5]

1. University of Washington, Department of Genome Sciences, Seattle, WA
2. University of Washington, Department of Pediatrics, Division of Genetic Medicine, Seattle, WA
3. University of Washington, Department of Laboratory Medicine and Pathology, Seattle, WA
4. Erasmus MC University Medical Center, Department of Clinical Genetics, Rotterdam, The Netherlands
5. University of Washington, Brotman Baty Institute for Precision Medicine, Seattle, WA

Chromatinopathies are rare neurodevelopmental disorders with signature epigenetic alterations (episignatures) caused by defects in chromatin modification machinery. Episignature-based classification has greatly increased the diagnostic yield, but current gold-standard array methods rely on bisulfite conversation, requiring additional sample manipulation, lengthening time to diagnosis, and interrogating limited genomic regions. Long-read sequencing (LRS) provides simultaneous whole-genome methylation and variant detection, potentially enabling rapid and robust diagnosis. Recent work has shown strong correlation between Oxford Nanopore LRS and Illumina EPIC array methylation calls, with successful classification of some affected individuals using publicly available array-derived episignatures.

To address these limitations, we developed EpiSignaLR, a streamlined classification tool integrating 34 established array-derived episignatures while enabling derivation of episignatures from LRS data. This dual approach is critical because current signatures do not perform reliably across all conditions. By leveraging the broader methylation coverage, EpiSignaLR provides a more complete view of the epigenome and resolves statistically significant differential methylation undetectable by arrays alone.

We applied EpiSignaLR to two chromatinopathies of differing analytical complexity: SETD1B-related disorder and Coffin-Siris syndrome (CSS). For SETD1B-related disorder, a LRS-derived episignature was robust and well-defined, revealing significantly more differentially methylated positions and novel methylated sites than array-based methods. CSS proved more challenging, reflecting the underlying genetic and epigenetic heterogeneity of BAF complex disorders, but a discriminating LRS-derived signature was nevertheless attainable. Together, these cases illustrate that LRS-derived episignatures can be defined across a spectrum of difficulty, including conditions where array-based approaches have been less successful.

We will further characterize these episignatures by evaluating classification performance across coverage thresholds to define the minimum depth required for clinical use. LRS-exclusive positions are anticipated to expand diagnostic resolution beyond array-based methods. These analyses demonstrate how methylation testing can be integrated into existing LRS clinical pipelines, reducing orthogonal testing and shortening the diagnostic odyssey for individuals with suspected chromatinopathies.

POSTER

143

Long-read transcriptome analysis using IsoRanker for identifying pathogenic variants in Mendelian conditions

Yong-Han Hank Cheng[1], Adriana E. Sedeño-Cortés[1], Jane E. Ranchalis[1], Katherine M Munson[1], Mitchell R. Vollger[1], Elsa Balton[1], Casie A. Genetti[2], Monica H. Wojcik[2], Alan H. Beggs[2], Michael J. Bamshad[1], Chia-Lin Wei[1], Katrina M. Dipple[1], Runjun D. Kumar[1], Elizabeth E. Blue[1], Ian Glass[1], Gail Jarvik[1], Jessica X. Chong[1], Daniela M. Witten[1], Anne O'Donnell-Luria[2], Andrew B. Stergachis[1]

1. University of Washington, Seattle, WA, USA
2. Boston Children's Hospital, Boston, MA, USA

Identifying pathogenic non-coding variants that contribute to Mendelian conditions remains challenging as the functional impact of these variants on gene function is often unknown. We present IsoRanker, a long-read transcriptome sequencing-based framework that prioritizes functionally relevant non-coding variants by detecting genes and novel isoforms with outlier expression, allelic imbalance, and/or nonsense-mediated decay (NMD). We generated paired cycloheximide-treated (NMD-inhibited) and untreated fibroblast transcriptomes from 31 individuals (3 individuals with known transcript-altering rare variants and 28 individuals with unsolved conditions) and linked transcripts to phased long-read genomes. IsoRanker successfully recovered known transcript alterations in this cohort and remained robust in subsampling analyses to cohorts of 11 individuals and ~5 million full-length transcripts per individual. However, performance was dependent upon de novo isoform caller choice, particularly for NMD-sensitive and novel isoforms. Among 28 previously unsolved cases, IsoRanker deprioritized most fibroblast-expressed candidate splice site variants while nominating new leads. In one individual, IsoRanker prioritized *HARS1*, revealing biallelic non-coding variants that together produced a partial *HARS1* loss-of-function and informed targeted therapy in this individual using histidine supplementation. These findings establish long-read, NMD-aware transcriptomics with IsoRanker as an effective approach for generating isoform-level functional evidence, improving classification of non-coding variants, and supporting the diagnosis of individuals with rare diseases.

POSTER

144

Machine learning-driven one health risk stratification of antimicrobial resistance and virulence in zoonotic pathogens across United States poultry, cattle, and human isolates

Emmanuel Kuufire[1], Viona Osei[1], Emmanuel Piiru[1], Rejoice Nyarku[1], Kingsley Bentum[1], Tyric James[1], Asmaa Elrefaey[1], Temesgen Samuel[1], Woubit Abebe[1]

1. Tuskegee University, Department of Pathobiology, Tuskegee, AL, USA

The growing burden of antimicrobial resistance (AMR) among foodborne zoonotic pathogens threatens both animal and human health and highlights the need for integrated One Health surveillance strategies. While genomic surveillance programs generate vast quantities of pathogen sequence data, most analyses remain descriptive and provide limited support for actionable risk prioritization. We developed a machine learning-enabled framework that integrates AMR determinants, virulence profiles, and cross-host genomic relatedness to support Precision Health risk stratification across poultry, cattle, and human reservoirs.

A total of 4,487 *Salmonella enterica* serovar Typhimurium and 3,820 *Campylobacter jejuni* isolates from the NCBI Pathogen Detection database (2015–2025) were analyzed. AMR genes were encoded as genomic features, virulence determinants were identified through comparative genome screening, and cross-host transmission potential was assessed through shared SNP clusters. Six machine learning algorithms were evaluated using stratified five-fold cross-validation to predict host reservoir, virulence burden, cross-host cluster membership, and high-risk resistance genotypes. An interpretable AMR-virulence scoring framework was developed to enable transparent risk classification.

We identified 71 *Salmonella* and 192 *Campylobacter* SNP clusters shared across multiple host sectors, indicating substantial genomic connectivity among animal and human reservoirs. Recurrent cross-host resistance determinants included *mdsA, mdsB, tet(A), sul2,* and *blaCMY-2* in *Salmonella*, and *tet(O), blaOXA-193,* and *gyrA-T86I* in *Campylobacter*. Host-reservoir classification achieved macro-F1 scores of 0.89 for *Salmonella* and 0.80 for *Campylobacter*. Integrated AMR-virulence scoring reclassified 2,293 multidrug-resistant *Salmonella* and 151 multidrug-resistant *Campylobacter* isolates into critical-risk categories that were not identified using AMR metrics alone. Prospective validation using post-2023 genomes demonstrated continued predictive performance for *Salmonella* (macro-F1 = 0.74) but not *Campylobacter* (macro-F1 = 0.26), consistent with NCBI convenience-sample bias.

This integrated workflow advances genomic surveillance beyond pathogen detection by combining machine learning, comparative genomics, and interpretable AMR-virulence scoring to prioritize high-risk zoonotic lineages and support surveillance, outbreak preparedness, and antimicrobial stewardship.

Mapping functional gene sets to cell populations using single-cell transcriptomics

POSTER

145

Hajar Amini[1], Colleen E. Clancy[1]

1. University of California, Davis, Sacramento, CA, USA

Determining which cell populations are primarily associated with a biological function, developmental program, or disease process is fundamental to understanding tissue organization and identifying potential therapeutic targets. Single-cell transcriptomics provides an unprecedented opportunity to characterize gene expression at cellular resolution, yet existing gene set scoring approaches evaluate cells independently and do not identify the coherent cell population that collectively represents a predefined functional gene set.

Here, we present a graph-based optimization framework for identifying the cell population most strongly associated with any predefined functional gene set from single-cell transcriptomic data. The method constructs a graph of transcriptionally similar cells and identifies a compact, connected cell population that jointly maximizes coordinated expression of the gene set, overall gene set expression within individual cells, and transcriptional similarity among selected cells. By prioritizing coherent cellular communities rather than isolated high-expressing cells, the framework maps functional gene sets to their dominant cellular context. An optional hierarchical search strategy enables efficient analysis of large-scale single-cell atlases while preserving single-cell resolution.

We applied the framework to ten curated cardiac functional gene sets representing normal physiology and cardiovascular disease, including cardiac conduction, muscle contraction, calcium handling, adrenergic signaling, angiogenesis, TGF-β signaling, cardiac development, and dilated and hypertrophic cardiomyopathy, using a human heart single-cell and single-nucleus transcriptomic atlas containing approximately 704,000 cells and nuclei across eight anatomical regions. The framework consistently identified a single dominant cell population for each gene set. Angiogenesis and VEGF signaling localized to endothelial cells, TGF-β signaling to fibroblasts, and muscle contraction, calcium handling, and the cardiomyopathy pathways to ventricular cardiomyocytes, cleanly separating the vascular, fibrotic, and contractile programs into their respective cell-type compartments.

Together, this framework provides a scalable and general approach for mapping predefined functional gene sets to the cell populations that most strongly coordinate their expression, enabling biological interpretation of large-scale single-cell datasets and prioritization of candidate cellular targets for mechanistic investigation.

POSTER

146

Mapping genetic effects on exercise response and associations with human traits using multi-omics and deep learning in the MoTrPAC human exercise cohort

Yilin Xie*[1], Esther Robb*[2], Sohaib Hassan[3], Salil Deshpande[4], Leanna M. Ross[5], Kristen Sutton[6], Nicolas Musi[7], William E. Kraus[5], Anshul Kundaje[2,8], Stephen B. Montgomery[1,2,3,8], MoTrPAC Study Group

1. Stanford University, Department of Pathology, Stanford, CA, USA
2. Stanford University, Department of Computer Science, Stanford, CA, USA
3. Stanford University, Department of Biomedical Data Science, Stanford, CA, USA
4. Stanford University, Institute for Computational and Mathematical Engineering, Stanford, CA, USA
5. Duke University, Department of Medicine, Durham, NC, USA
6. University of Colorado, Department of Biomedical Informatics, Aurora, CO, USA
7. Cedars-Sinai Medical Center, Los Angeles, CA, USA
8. Stanford University, Department of Genetics, Stanford, CA, USA

*Co-first Author

Physical exercise has broad benefits on human health and physical performance, and is associated with lower all-cause mortality and reduced risk of cardiovascular disease and some cancers. But our understanding of how genetics influence exercise response and health outcomes is limited. The Molecular Transducers of Physical Activity Consortium (MoTrPAC) has generated genomic and multi-omics data from 174 sedentary adult participants during and after acute endurance (~40 min cycling) and resistance exercise, sampled at timepoints ranging from 20 min into exercise to 24 hr post-exercise, providing a unique opportunity to link genetics to exercise-induced molecular responses. Leveraging this dataset, we analyzed the interaction of genetics and exercise on gene regulation and expression and mapped the relationship with diseases. We integrated 72 human traits GWASs through stratified LD score regression (S-LDSC) and transcriptome-wide association study and found that exercise molecular response is enriched for broad disease risk heritability, especially autoimmune traits and cardiometabolic disease. We then developed a response QTL model capturing timepoint-specific molecular effects of endurance and resistance exercise, which discovered 249 exercise-dependent eQTL genes and 41 differential allele specific expression effects in blood, involving genetic risk variants associated with human phenotypes, including lipid metabolism and anthropometric traits.

We subsequently developed a novel deep-learning based variant effect predictor to survey exercise-dependent variant effects, enabling us to extend our analysis to power-limited settings. We leveraged the ChromBPNet model on ATAC-seq from skeletal muscle tissue in samples from control, resistance, and endurance groups to identify variants with differential effects on chromatin accessibility. This method nominated 4462 exercise-dependent variants among common (MAF>1%) variants in the MoTrPAC human cohort, of which 1084 overlapped with GWAS finemapped trait loci. Using S-LDSC, exercise-dependent variants showed significant heritability enrichment for cardiometabolic traits, including heart attack and type 2 diabetes. Among exercise-dependent variants we found 2153 that altered known transcription factor motifs, and we validated differential motifs in phosphoproteomics data. Combined, our analysis identifies the molecular mechanisms of gene-by-exercise interactions, and demonstrates an AI-based framework for identifying gene-by-exercise variants that we apply to enable genome-wide and allele-frequency independent analyses of the relationship between genetics, exercise, health and disease.

POSTER

147

Modeling sequencing depth requirements for clinical long-read whole-genome somatic variant detection

Andrew Todeschini[1], Ryan Miller[1], DianRong Tsai[1], Ken Dixon[2], Jared Barrott[1,2,3,4]

1. Brigham Young University, Cell Biology and Physiology, Provo, UT, USA
2. SpeciCare, Gainesville, GA, USA
3. Brigham Young University, Simmons Center for Cancer Research, Provo, UT, USA
4. Brigham Young University, School of Medicine, Provo, UT, USA

Background: Long-read whole-genome sequencing (WGS) is increasingly being adopted for cancer genomics due to its ability to resolve complex genomic variation. However, the sequencing depth required to achieve reliable somatic variant detection while minimizing cost remains poorly defined. We evaluated the effect of sequencing depth on somatic mutation recovery to identify practical coverage thresholds for clinical nanopore WGS.

Methods: Matched Oxford Nanopore Technologies (ONT) and Illumina WGS datasets from eight solid tumor specimens were analyzed. ONT sequencing data were computationally downsampled to 30×, 15×, 10×, and 5× coverage and compared with matched Illumina-derived somatic variant calls. In parallel, Illumina variant calls were computationally downsampled using identical variant-calling thresholds to distinguish losses attributable to sequencing depth from platform-specific performance. Variant concordance, variant allele frequency (VAF), total read depth, and mutant allele counts were evaluated across sequencing depths. Structural variant concordance between platforms was also assessed.

Results: Somatic variant concordance declined progressively with decreasing ONT coverage, with the greatest reduction observed at 5× depth. Variants missed at lower sequencing depths were consistently associated with lower VAFs and reduced mutant allele support, whereas total read depth at variant loci remained relatively stable across coverage levels. Similar patterns observed following Illumina downsampling indicate that reduced recovery of low-frequency variants is driven primarily by sampling limitations rather than platform-specific sequencing error. Structural variant overlap between ONT and Illumina varied across tumor samples, demonstrating complementary detection capabilities for larger genomic alterations.

Conclusions: Sequencing depth substantially influences somatic mutation sensitivity in long-read WGS, with low-VAF variants representing the principal source of detection loss at reduced coverage. These findings provide an empirical framework for selecting sequencing depth in clinical nanopore cancer sequencing and support coverage optimization strategies that balance analytical sensitivity with sequencing cost. Future predictive models incorporating variant allele frequency and mutant allele support may further refine depth recommendations for precision oncology applications.

Mutagenesis sensitivity mapping of human enhancers in vivo

POSTER

148

Michael Kosicki[1], Boyang Zhang[2], Anusri Pampari[2], Jennifer A. Akiyama[1], Ingrid Plajzer-Frick[1], Catherine S. Novak[1], Stella Tran[1], Yiwen Zhu[1], Momoe Kato[1], Riana D. Hunter[1], Kianna von Maydell[1], Sarah Barton[1], Erik Beckman[1], Anshul Kundaje[2], Diane E. Dickel[1], Axel Visel[1], Len A. Pennacchio[1]

1. Lawrence Berkeley National Lab, Berkeley, CA, USA
2. Stanford University, Department of Computer Science, Stanford, CA, USA

Mutations in regulatory DNA account for a large fraction of variants suspected to contribute to human diseases. Their interpretation requires better models of in vivo enhancer logic. Machine learning models promise to advance our understanding of regulatory DNA activity, but they require systematic mutagenesis of enhancers in vivo for validation. Such studies have been hindered by the lack of high-throughput tools to interrogate sequence perturbations.

To address these challenges, we systematically mutagenized seven enhancers that reliably promote tissue-specific expression in developing mouse embryos. We applied our highly specific and reproducible transgenic enSERT strategy to assess the in vivo consequences of these mutations, creating over 1700 transgenic embryos for more than 260 mutagenized constructs. These results were then compared to predictions by ChromBPNet models trained on embryonic, tissue-specific open chromatin datasets.

Our analyses revealed a diversity of functional site arrangements within enhancers, complementary modes of multi-tissue activity, and complexity of activator-repressor logic. They also demonstrated high positive predictive value and considerable sensitivity of machine learning models for prediction of functional binding motifs at base-pair resolution. These results underscore the complexity of regulatory elements and provide a foundation for further development of computational approaches for regulatory variant prediction.

POSTER

149

Native long-read methylation sequencing identifies ovarian cancer-specific epigenetic signatures for precision diagnostics

DianRong Tsai[1], Ken Dixon[2], Jared Barrott[1,2,3,4]

1. Brigham Young University, Cell Biology and Physiology, Provo, UT, USA
2. SpeciCare, Gainesville, GA, USA
3. Brigham Young University, Simmons Center for Cancer Research, Provo, UT, USA
4. Brigham Young University, School of Medicine, Provo, UT, USA

Background: Native long-read methylation sequencing enables simultaneous measurement of DNA sequence and epigenetic modifications without bisulfite conversion, providing a scalable platform for precision biomarker discovery. Unlike conventional bisulfite sequencing, nanopore sequencing directly detects DNA methylation while preserving native DNA and long-range genomic context. We evaluated whether genome-wide native methylation profiling could distinguish ovarian cancer from normal ovarian tissue and other malignancies to support development of precision diagnostic biomarkers.

Methods: Genome-wide native DNA methylation profiles were generated using Oxford Nanopore Technologies (ONT) sequencing from cryopreserved ovarian cancer, normal ovarian tissue, and non-ovarian cancer samples. Sequencing detected approximately 28.3–28.9 million CpG sites per sample, with average CpG coverage ranging from 7.8× to 30.3×. All samples exceeded the predefined 5× minimum coverage threshold for downstream analysis, and coverage-based filtering excluded low-confidence CpG calls prior to generating the integrated methylation matrix. Global methylation profiles were evaluated using unsupervised principal component analysis (PCA). Ovarian cancer-specific differentially methylated regions (DMRs) are being identified and will be independently evaluated using orthogonal methylation datasets from The Cancer Genome Atlas (TCGA).

Results: Native ONT sequencing generated consistent genome-scale methylation profiles across all samples despite variability in sequencing depth. Samples with lower average coverage (7.8× and 12.0×) remained suitable for analysis following quality filtering, although additional low-confidence CpG sites were excluded. Unsupervised PCA demonstrated clear separation of ovarian cancer, normal ovarian tissue, and non-ovarian cancers, indicating that global methylation landscapes are highly tissue- and disease-specific. These findings demonstrate that native long-read methylation sequencing captures biologically meaningful epigenetic signatures capable of distinguishing ovarian cancer from both normal tissue and other malignancies without bisulfite conversion.

Conclusions: Native long-read methylation sequencing provides robust genome-wide epigenetic profiling with broad CpG representation across diverse tissue samples. Integration of ONT methylation profiling with orthogonal TCGA validation establishes a scalable framework for identifying reproducible ovarian cancer methylation biomarkers and supports future development of minimally invasive precision diagnostics and blood-based early detection assays.

Non-toxic HSV-1-mediated NaV1.7 knockdown for the targeted treatment of chronic neuropathic pain

POSTER

150

Nazanin Mahinparvar[1], Michael Klukinov[1], Phillip Kyriakakis[2], David C. Yeomans[1]

1. Stanford University School of Medicine, Department of Anesthesiology, Perioperative and Pain Medicine, Stanford, CA, USA
2. Stanford University, Department of Bioengineering, Stanford, CA, USA

Background: Chronic neuropathic pain remains difficult to treat because available therapies often provide incomplete relief and cause dose-limiting adverse effects. The voltage-gated sodium channel NaV1.7, highly expressed in dorsal root ganglion (DRG) nociceptors, is a validated target for pain modulation. Recombinant herpes simplex virus type 1 (HSV-1) vectors undergo retrograde transport to the DRG after peripheral administration, enabling targeted gene delivery to sensory neurons. We evaluated a single-vector, replication-defective HSV-1 platform encoding short hairpin RNA (shRNA) to suppress NaV1.7 in DRG neurons.

Methods: A replication-defective HSV-1 vector encoding NaV1.7-targeting shRNA and an mCherry reporter was generated using a BAC-based platform. In vitro assays assessed reporter expression, vector transduction, and cytotoxicity. Adult rats received intradermal hindpaw injections of HSV-NaV1.7, scrambled-shRNA HSV, or vehicle. Lumbar (L4-L6) and thoracic DRGs were harvested for blinded immunohistochemical and immunofluorescence analyses of vector distribution and NaV1.7 expression. Safety was evaluated by histological examination and clinical observation. Behavioral efficacy testing using von Frey and Hargreaves assays is ongoing.

Results: In vitro studies demonstrated robust mCherry expression and efficient vector transduction without detectable cytotoxicity. Following hindpaw administration, mCherry fluorescence was detected in L4-L6 DRG neurons but not in thoracic DRG controls, supporting anatomically localized retrograde transport. Quantitative immunohistochemical and immunofluorescence analyses demonstrated reduced NaV1.7 protein expression in HSV-NaV1.7-treated DRG neurons compared with both scrambled-shRNA and vehicle controls. Histological evaluation revealed no evidence of neuroinflammation, tissue pathology, or adverse local effects.

Conclusion: A replication-defective HSV-1 vector delivered NaV1.7-targeting shRNA to lumbar DRG neurons through retrograde transport and produced localized NaV1.7 suppression without detected cytotoxicity or tissue pathology. These findings establish target engagement and preliminary safety, supporting continued development of this platform as a localized, non-opioid gene therapy for chronic neuropathic pain. Behavioral studies are underway to determine analgesic efficacy.

Novel variants and candidate gene discovery for hearing loss using exome sequencing in Malian populations

POSTER

151

Abdoulaye Yalcouye[1,2,3], Isabelle Schrauwen[4], Oumou Traoré[1], Cheick O. Guinto[1,5], Guida Landouré[1,5], Suzanne M. Leal[4], Ambroise Wonkam[2,3]

1. Université des Sciences, des Techniques et des Technologies de Bamako (UTTTB), Faculté de Médecine et d'Odontostomatologie, Bamako, Mali
2. University of Cape Town, Faculty of Health Sciences, Department of Medicine, Division of Human Genetics, Cape Town, South Africa
3. Johns Hopkins University, School of Medicine, McKusick-Nathans Institute and Department of Genetic Medicine, Baltimore, MD, USA
4. Columbia University Medical Center, Center for Statistical Genetics, Gertrude H. Sergievsky Center, Department of Neurology, New York, NY, USA
5. Centre Hospitalier Universitaire du Point G, Service de Neurologie, Bamako, Mali

Hearing loss (HL) is the most common sensory disorder worldwide and one of the most genetically heterogeneous human conditions. Approximately half of congenital deafness is genetic, and genetic susceptibility also affects acquired forms of HL caused by noise exposure, ototoxic antibiotics, chemotherapy, and aging. More than 500 genes are predicted to contribute to syndromic and non-syndromic HL, yet only a subset has been identified. Gene discovery has slowed at precisely the moment when the field is entering a therapeutic era: the first gene therapies for HL are advancing rapidly, with recent FDA approval for OTOF-related HL. African populations provide an extraordinary and still largely untapped opportunity to transform HL genetics. Despite that, there is limited genetic data on HL from sub-Saharan African populations. In this study, we investigated the genetic causes of HL in Malian populations (West Africa) using whole-exome sequencing. Furthermore, cDNA was transfected into HEK293T cells for localization and expression analysis in a candidate gene.

Twenty-four multiplex families were enrolled, of which 50% (12/24) are consanguineous. Variants were found in six known non-syndromic HL (NSHL) genes, four genes that can underlie either syndromic HL (SHL) or NSHL, one SHL gene, and one novel candidate HL gene. Overall, 75% of families (18/24) were solved, and 94.4% (17/18) had variants in known HL genes, including MYO15A, CDH23, MYO7A, GJB2, SLC26A4, PJVK, OTOGL, TMC1, CIB2, GAS2, PDCH15, and EYA1. A digenic inheritance (CDH23 and PDCH15) was found in one family. Most variants, 59.1% (13/22) in known HL genes, were not previously reported or associated with HL. The UBFD1 candidate HL gene, which was identified in one consanguineous family, is expressed in human inner ear organoids. Cell-based experiments in HEK293T showed that mutants of UBFD1 had a lower expression compared to the wild type.

Together, these findings substantially expand the known genetic architecture of HL in an underrepresented African population and identify a large proportion of previously unreported pathogenic variants. We report the profile of known genes and the UBFD1 candidate gene for HL in Mali and emphasize the potential of gene discovery in African populations. This effort is timely since precision medicine for HL is rapidly evolving.

Keywords: Hearing Loss; Novel Variants; Candidate Gene; Precision Medicine; Mali; Africa.

PAG-Agent: a biologist-oriented virtual research assistant for context-aware pathway analysis, prioritization, and interpretation

POSTER

152

Quang-Huy Nguyen[1,2], Zeru Zhang[5], Zhandos Sembay[3], Duc-Hau Le[2], Jake Y. Chen[3], Wei-Shinn Ku[1], Hao Chen[4], Zongliang Yue[5]

1. Auburn University, Department of Computer Science and Software Engineering, Auburn, AL, USA
2. Hanoi University of Science and Technology, School of Information and Communications Technology, Hanoi, Vietnam
3. University of Alabama at Birmingham, Department of Biomedical Informatics and Data Science, Birmingham, AL, USA
4. Auburn University, College of Forestry and Wildlife Sciences, Auburn, AL, USA
5. Auburn University, Department of Health Outcomes Research and Policy, Auburn, AL, USA

Transcriptomic studies have become fundamental to disease research and precision health, yet translating differential expression results into biological insight remains challenging. First, conventional pathway analysis methods primarily rank pathways by statistical significance, which does not necessarily reflect their biological relevance to the experimental context or research objective. Second, transcriptomic analysis workflows remain fragmented, requiring researchers to switch between multiple tools for differential expression analysis, pathway analysis, biological annotation, literature validation, and hypothesis generation. We present PAG-Agent, a biologist-oriented virtual research assistant that addresses both challenges by integrating transcriptomic analysis and post-analysis within a unified platform while introducing context-aware Over-Representation Analysis (cORA) for biologically informed pathway prioritization. Unlike conventional enrichment methods that perform pathway analysis on an entire gene set, cORA first constructs a confidence-weighted gene interaction network from HAPPI2 and identifies overlapping co-expressed gene modules using oCEM, allowing genes to participate in multiple functional modules and better reflecting the pleiotropic nature of biological systems. It then performs ORA on individual modules before prioritizing enriched pathways according to module quality, experimental conditions, and study objectives. Beyond pathway prioritization, PAG-Agent provides an end-to-end transcriptomic workflow through natural language interaction, eliminating the need to switch between independent tools for differential expression analysis, pathway analysis, biological annotation, literature validation, and hypothesis generation. We evaluated cORA on 60 independent bulk and single-cell transcriptomic studies spanning diverse biological conditions. Using study-reported pathways as context-specific references, cORA consistently outperformed conventional ORA and eight large language model-based approaches. At a semantic similarity threshold of 0.5, cORA achieved an accuracy of 0.913, a hit rate of 0.902, and an F1 score of 0.898. Ablation analyses demonstrated that both biologically informed module detection and contextual prioritization contributed substantially to performance. We further evaluated PAG-Agent on three Alzheimer's disease transcriptomic datasets, where it supported the complete transcriptomic analysis workflow from differential expression analysis to biological interpretation and consistently identified neurodegeneration-associated pathways across multiple analytical methods and datasets. Taken together, PAG-Agent combines biologically informed pathway prioritization with an integrated transcriptomic analysis workflow, enabling researchers to translate transcriptomic measurements into context-specific, literature-supported biological hypotheses for precision health research.

Patient-to-patient phenotype matching enables rare disease gene prioritization beyond curated gene–phenotype associations

POSTER

153

Isabelle Cooperstein[1], Alistair Ward[1,2], Shilpa N. Kobren[3], Emerson Lebleu[1], Barry Moore[1], Rebecca C. Spillmann[4], Vandana Shashi[4], Undiagnosed Diseases Network, Gabor T. Marth[1]

1. University of Utah, Department of Human Genetics, Salt Lake City, UT, USA
2. Frameshift Labs Inc, Cambridge, MA
3. Harvard Medical School, Department of Biomedical Informatics, Boston, MA
4. Duke University School of Medicine, Division of Medical Genetics, Department of Pediatrics, Durham, NC

Existing phenotype-aware prioritization methods infer diagnoses through curated disease models and predefined gene associations, limiting performance for atypical presentations, newly emerging disorders, and genes with sparse or absent phenotype associations. SimPheny, a phenotype-first patient matching method, instead derives diagnostic evidence directly from phenotypic similarity between real patients, who better capture the phenotypic diversity and diagnostic complexity of rare conditions.

SimPheny introduces a two-stage prioritization strategy. First, pairwise phenotypic similarity is computed between a query patient and diagnosed reference patients using Human Phenotype Ontology (HPO)-based semantic similarity. Second, candidate genes are prioritized by evaluating overlap between a patient's candidate gene list and the causal genes of phenotypically similar diagnosed cases. Each overlap defines a SimPheny Match that integrates phenotypic similarity significance with gene match specificity to generate a quantitative diagnostic score and confidence classification.

We benchmarked SimPheny using diagnosed probands from the Undiagnosed Diseases Network (UDN), then applied it to a discovery cohort of unsolved UDN cases. To expand match sensitivity, we increased the size and heterogeneity of the diagnosed reference cohort using solved cases from ClinVar, DECIPHER, and the Phenopacket Store.

SimPheny outperformed existing phenotype-aware tools across diagnosed UDN probands, never placing the diagnostic gene beyond the 15th candidate, whereas many comparison tools extended beyond 50 candidates. For genes with limited or no phenotype associations, SimPheny consistently recovered the diagnostic gene within the top two candidates, demonstrating its ability to detect diagnoses that association-dependent tools miss. In unsolved UDN probands, SimPheny identified 18 high-confidence matches, of which 8 (44.4%) were ultimately confirmed following review. Diagnostic ranking performance remained robust despite substantial expansion and increased heterogeneity of the diagnosed reference cohort, with additional matches from external datasets strengthening support for unresolved candidate diagnoses.

These findings establish patient-to-patient matching as a scalable strategy for rare disease gene prioritization, supporting reanalysis and earlier recognition of emerging gene-disease relationships as new diagnosed cases become available.

POSTER

154

Performance benefits and limits of fine-tuning sequence-to-function models on biobank-scale cohorts

Eduarda Vaz[1], Lena Wang[2], Jake Galvin[3], Rebecca Keener[4], Alexis Battle[4,5,6,7,8]

1. Johns Hopkins University, Department of Chemical and Biomolecular Engineering, Baltimore, MD, USA
2. Johns Hopkins University, Department of Molecular and Cellular Biology, Baltimore, MD, USA
3. Johns Hopkins University, Cell, Molecular, Developmental Biology, and Biophysics Program, Baltimore, MD, USA
4. Johns Hopkins University, Department of Biomedical Engineering, Baltimore, MD, USA
5. Johns Hopkins University, Department of Computer Science, Baltimore, MD, USA
6. Johns Hopkins University, Department of Genetic Medicine, Baltimore, MD, USA
7. Johns Hopkins University, Malone Center for Engineering in Healthcare, Baltimore, MD, USA
8. Johns Hopkins University, Data Science and AI Institute, Baltimore, MD, USA

Accurate prediction of the molecular consequences of sequence variation is a central goal of human genetics. Sequence-to-function models predict regulatory activity across the human genome but struggle to predict the regulatory impact of individual genetic variation. Here, we evaluated whether model performance could be improved by incorporating biobank-scale cohorts with paired whole-genome sequencing (WGS) and functional data, moving beyond the human reference genome or modest cohort sizes.

We fine-tuned Borzoi models using WGS and plasma proteomics from 54,219 UK Biobank individuals and evaluated generalization across held-out individuals or genes. First, we tested 400 models each fine-tuned for a single gene and demonstrated that fine-tuning improved personal protein level prediction on held-out individuals for 80% of genes relative to a linear baseline. We asked whether these gains reflected additive genetic effects or more complex functions of personal genetic sequence. While most genes were well explained by a dosage-weighted additive approximation, highly predictable genes showed systematic deviations from additive models. Genotype-aware and pairwise-interaction models improved agreement with fine-tuned Borzoi, suggesting nonlinear effects contribute to prediction for some genes. Training sample size was also a major determinant of model performance. Full-cohort models improved accuracy over models trained on 1,000 individuals by $\Delta R2 = 0.1$, recovered 67% more SuSiE fine-mapped pQTLs (PIP > 0.9) at the top 1% saliency threshold and showed stronger agreement between predicted variant effects and observed pQTL β. Gains were also associated with higher prioritization of variants with putative functional evidence and reduced LD redundancy. We further explored models fine-tuned across many genes, evaluating their performance on genes excluded from training as a stringent test of generalization to sequence variation at unseen loci. Even at sample sizes of tens of thousands of individuals, gains over baseline were evident but modest. Together, these results suggest that single-gene models benefit from larger cohorts by learning nonlinear effects and low-frequency QTLs within each locus. However, consistent with recent work, the fine-tuned models primarily learn observed genetic associations rather than general principles of functional regulation. Our findings highlight the need to complement large datasets with improved training strategies and benchmarks.

POSTER

155

Platform-driven discovery of antisense oligonucleotide therapeutics for dominant haploinsufficiency

Stephen Tran[1], Kun Tan[2], Katherine Rothamel[3], Lilia Iakoucheva[1], Miles Wilkinson[2], Gene Yeo[3], Jonathan Sebat[1,3]*

1. University of California, San Diego, Department of Psychiatry, San Diego, CA, USA
2. University of California, San Diego, Department of Obstetrics, Gynecology, and Reproductive Sciences, San Diego, CA, USA
3. University of California, San Diego, Department of Cellular and Molecular Medicine, San Diego, CA, USA

Dominant haploinsufficiency is a major cause of severe neurodevelopmental disorders (NDDs), yet therapeutic strategies to restore gene expression remain limited. Because one functional allele remains, increasing expression from the intact copy represents an attractive therapeutic strategy. One promising approach uses splice-modulating antisense oligonucleotides (ASOs) targeting poison exons (PEs), naturally occurring alternatively spliced exons that reduce productive mRNA output by introducing premature termination codons and triggering nonsense-mediated decay (NMD). ASOs can inhibit PE inclusion and thereby increase gene expression, a strategy that has recently entered clinical validation studies for SCN1A-associated Dravet syndrome with promising early results.

Despite this progress, broader application of PE-targeting therapeutics has been limited by incomplete knowledge of PE biology and uncertainty regarding which PEs function as therapeutically actionable regulators of gene expression. To address this challenge, we developed a transcriptome-wide platform for discovery and prioritization of PE-ASO targets by integrating deep PacBio long-read sequencing of NMD-deficient human H9 embryonic stem cells with transcriptomic data from 30 NMD-perturbed cellular models spanning embryonic stem cells, neural precursor cells, neurons, and laboratory cell lines.

Our platform identified 28,181 NMD-enriched transcript isoforms, 82% of which were previously unannotated, revealing multiple classes of poison exons and other NMD-associated splicing events. Meta-analysis across the 30 datasets identified 10,819 reproducibly NMD-regulated poison exons and nominated candidate PE-ASO targets in 141 dominant haploinsufficiency genes.

To evaluate the predictive performance of the platform, we selected top-ranked candidates for ASO development. For DHX9 and HNRNPL, 2′-MOE-modified ASOs targeting PE splice sites reduced PE inclusion by ~99% and increased total mRNA levels by 1.5–2-fold in HEK293T cells. Validation of additional prioritized targets is ongoing.

Together, these results establish a scalable framework for transcriptome-wide discovery and prioritization of therapeutically actionable poison exons and provide a foundation for precision ASO development across dominant haploinsufficiency disorders.

POISEN: a tool for poison exon discovery and gene-targeting therapy opportunities

POSTER

156

Mia S. Broad[1], **Sheng Tang**[1], **Namitha Alexander**[1], **Jung Hong**[1], **Jeffrey D. Calhoun**[1], **Gemma L. Carvill**[1]

1. Northwestern University, Neurology, Chicago, IL, USA

Alternative poison exon (PE) splicing regulates protein abundance with cell-type specificity. When included in mRNA transcripts, PEs introduce premature termination codons that trigger nonsense-mediated mRNA decay (NMD). Aberrant PE splicing has been implicated in neurological disorders and cancers. For example, we identified rare pathogenic *SCN1A* variants in individuals with Dravet syndrome that cause aberrant PE splicing in iPSC-derived neurons, leading to increased PE inclusion and likely reduced protein abundance consistent with *SCN1A* haploinsufficiency. However, PEs can also be therapeutic targets: antisense oligonucleotides (ASOs) that prevent PE inclusion can rescue haploinsufficiency, and an *SCN1A*-PE-targeting ASO has shown up to an 80% reduction in seizures in patients with Dravet syndrome in phase 1 clinical trials.

PEs may therefore help explain missing genetic etiology and serve as targets for precision therapeutics. Yet PEs remain understudied due to biological and technological challenges. PE-containing isoforms are rapidly degraded by NMD, limiting detection, and short-read RNA sequencing cannot resolve splice junction order due to read length limitations. Consequently, PEs have been characterized from a limited set of well-studied examples as highly conserved, short (<50–100 bp) exons that occur in approximately 30% of the human transcriptome.

To address these limitations, we developed POISEN (Poison exOn dIScovery for long-rEad traNscriptomes), a first-in-class, PE-specific bioinformatics pipeline that identifies and classifies PEs from long-read transcriptomes and validates them using NMD inhibition. In HepG2 and HAP1 cells, POISEN discovered 7,817 validated PEs, including 5,196 novel PEs (66.5%) not previously annotated in GENCODE v48. Approximately 24% of expressed genes contained at least one PE (n = 4,200 genes), and 1,933 PE-containing genes encoded two or more PEs. Nearly 75% of validated PEs were >100 bp, with some extending to 1–5 kb. Additionally, PEs showed intermediate evolutionary constraint, with a median GERP score (0.174) lower than coding exons (0.978) but higher than intronic (-0.267) and enhancer (-0.187) regions.

Among genes harboring validated PEs, 1,301 of 4,200 genes (31%) were disease-associated in GenCC, representing significant enrichment (OR = 1.45, p = 3.66e-20). Collectively, POISEN provides a novel framework for defining PE-mediated regulation and identifying candidate therapeutic ASO targets for monogenic disorders.

POSTER

157

Population-scale long-read sequencing reveals molecular and clinical consequences of complex genetic variation in All of Us

Qiuhui Li[1], Fabio Cunial[2], Hang Su[2], Kiran V. Garimella[2], Julie Wertz[3], Samuel K. Lee[2], Yongqing Huang[2], Yulia Mostovoy[2,4], Adam English[5], Donna M. Muzny[5], Matt C. Danzi[6], Mary Wons[2], Yuanyuan Wang[7], Isaac Wong[3], Isaac R. L. Xu[6], Evin Padhi[8], Stephan Zuchner[6], Lee Lichtenstein[2], Namrata Gupta[2,4], Niall Lennon[2,4], Stacey Gabriel[2,4], Jane Grimwood[9], Stephen B. Montgomery[8], Kimberly F. Doheny[1], Tara Dutka[10], Anjene Musick[10], Fritz J. Sedlazeck[5], Michael C. Schatz[1], Evan E. Eichler[3], Michael E. Talkowski[2,4]

1. Johns Hopkins University, Baltimore, MD, USA
2. Broad Institute of MIT & Harvard, Cambridge, MA, USA
3. University of Washington, Seattle, WA, USA
4. Massachusetts General Hospital, Boston, MA, USA
5. Baylor College of Medicine, Houston, TX, USA
6. University of Miami, Miami, FL, USA
7. Vanderbilt University, Nashville, TN, USA
8. Stanford University, Stanford, CA, USA
9. HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA
10. National Institutes of Health (NIH), Bethesda, MD, USA

Structural variants (SVs), tandem repeats, mobile element insertions, and other forms of complex genetic variation contribute to human phenotypic diversity but remain difficult to resolve with short-read sequencing. Here, we generated PacBio HiFi whole-genome sequencing data for 12,212 participants in the All of Us (AoU) Research Program, spanning six continental ancestry groups with more than 80% participants of non-European ancestry. Participants were sequenced at 15× (n=11,079) or 30× (n=1,133) coverage. Matched RNA-seq (n=6,451), Olink proteomics (n=8,822), CpG methylation data (n=12,212), and longitudinal electronic health records (n=10,517; 9,370 with ICD codes) enabled molecular and clinical analyses of complex variation.

We constructed an integrated, haplotype-resolved callset comprising SNVs, indels, 2.5M SVs ≥50 bp, and tandem-repeat genotypes at 4.4M loci. The callset incorporated the AoU participants together with 292 HPRC and HGSVC samples with reference diploid assemblies for quality assessment. Association testing across thousands of EHR-derived phenotypes identified signals involving HLA alleles, KIR haplotypes, mitochondrial haplogroups and heteroplasmies, and mobile element insertions. Novel immunogenetic associations included HLA-DRB1*15:02 with increased weight and waist-to-hip measurements ($p<10^{-12}$), HLA-B*38:01 and HLA-DPB1*18 with lower heart rate, and KIR3DL2*007 with acute posthemorrhagic anemia, potentially reflecting allele-specific effects on natural killer cell function.

We systematically characterized the molecular consequences of SVs, identifying an extensive landscape of SV expression, splicing, and protein quantitative trait loci (eQTLs, sQTLs, and pQTLs). Joint fine-mapping with nearby SNVs supported SVs as candidate causal variants at thousands of molecular-trait loci and prioritized them as the top variants at hundreds of loci. Molecular-outlier analyses further detected over 2,000 rare regulatory SVs associated with extreme gene expression or protein abundance that are inaccessible to conventional QTL mapping. Phenome-wide association studies (PheWAS) identified more than 900 significant SV-disease associations, with prioritized associations evaluated in an independent, ancestry-matched short-read AoU cohort of ~100,000 participants. Colocalization analyses linked molecular QTLs with disease-association signals across many conditions including respiratory malignancy and hypothyroidism, while transcriptome-wide association studies (TWAS) prioritized candidate effector genes. Together, these results demonstrate how population-scale long-read sequencing connects complex genetic variation with molecular function and human disease across diverse populations.

POSTER

158

Practical utility of sequence-to-omics models for improving the reproducibility of genetic fine-mapping

Michael D. Sweeney[1], **Hyun Min Kang**[1]

1. University of Michigan, Department of Biostatistics and Center for Statistical Genetics, Ann Arbor, MI, USA

Deep learning- and AI-based sequence-to-omics (S2O) models—such as AlphaGenome, Borzoi, Enformer, and Sei—predict molecular phenotypes directly from DNA sequence, offering a route to prioritize functional variants orthogonal to statistical association. Statistical fine-mapping methods are the current standard approach for pinpointing putative causal variants, yet often replicate poorly across independent studies due to factors like linkage disequilibrium, imputation uncertainty, and rare causal variants. Here, we evaluate to what extent S2O models can improve the reproducibility of genetic fine-mapping across expression quantitative trait loci (eQTL) datasets from GTEx, TOPMed, and MAGE.

Using a discovery-replication framework, we found that, compared to statistical posterior inclusion probabilities (PIPs), S2O model scores improved inter-study replication rates (ISRRs) among high-confidence singleton (PIP>95%) GTEx blood eQTL credible sets in TOPMed. Across the four models, SNPs in the highest decile of S2O scores achieved an ISRR of 86.4%-91.0%, whereas SNPs in the highest statistical PIP decile showed a modest ISRR of 78.3%.

We developed a generalizable framework for functionally informed fine-mapping that yields functionally informed PIPs (fiPIPs) by reweighting PIPs using variant-to-probability scores from an XGBoost leave-one-chromosome-out classifier trained on S2O scores. Among GTEx blood credible sets of size two with exactly one replicated SNP, fiPIPs prioritized the replicated SNP up to 79.3% of the time compared to 66.6% for PIPs. We noticed a similar trend with GTEx blood credible sets of size 3-10 SNPs; fiPIPs prioritized a replicated SNP up to 70.0% of the time compared to 49.9% for PIPs. We noticed similar trends with lymphoblastoid cell line (LCL) eQTLs from GTEx/MAGE. We also found that, compared to PIPs, fiPIPs engendered higher Spearman correlations with published massively parallel reporter assay (MPRA) allelic skew magnitudes.

Our findings demonstrate that S2O models, especially newer architectures like AlphaGenome and Borzoi, substantially improve identification of reproducible SNPs across studies. Our released framework facilitates scalable, functionally aware fine-mapping by flexibly using any quantitative scores, so it can be used as AI technologies advance. By adding an orthogonal, functional signal that resolves SNPs that statistics alone cannot, S2O-informed fine-mapping delivers a reliable and interpretable foundation for translating genomic discoveries into clinical insight.

Precision medicine requires novel tools to breach the widespread clinical adoption barrier

POSTER

159

Wilfrido Mojica[1], Alexander Arrow[1], Emily Chen[2], Keith Booher[2], Natalie Lutala[3], Chang-Chieh Hsu[4], Yun Wu[4], Troy Wood[5], Alexandra M Izydorczak [6], Donald Yergeau[7]

1. Virchow Medical, Inc., La Jolla, CA, USA
2. Zymo Research, Tustin, CA, USA
3. Crown Bioscience, San Diego, CA, USA
4. University at Buffalo, Department of Biomedical Engineering, Buffalo, NY, USA
5. University at Buffalo, Department of Chemistry, Buffalo, NY, USA
6. University of Pittsburgh, Department of Psychiatry, Pittsburgh, PA, USA
7. University at Buffalo, Genomics and Bioinformatics Core, Buffalo, NY, USA

Precision Medicine is rapidly expanding and encompasses not only genomics, but emerging fields such as organoids, metabolomics and extracellular vesicle research. Although tools for genomics have been able to bridge the barrier of formalin-fixed, paraffin embedded processing that is used by laboratories to make clinical samples accessible, other technologies remain blocked. This delays and may even prevent the introduction of these promising diagnostic methodologies into clinical practice. An approach capable of recovering derivable elements from diagnostic tissue specimens without compromising it for downstream morphologic evaluation needs to be developed in order to bring the full armamentarium of emerging tools becoming available into the battle against cancer. Virchow Medical has developed and is continuing to create a suite of tools that addresses this problem and spans the breadth of clinical specimen types from cytology-based methods (the Cell Wrangler), to small biopsy procedures (the Crow's Nest) and extending to definitive resections (the MOOS). Developed by a pathologist with an eye towards preservation of the tissue and minimizing the critical questions that may arise from morphology-based end-users, these tools share several key features that include the intention for multi-analytics and the absence of exposure to formalin. The potential application of material captured by these tools is enormous and limited only by resources and intent. From material recovered by these tools we have been able to document the ability to grow organoids, recover DNA in the 100-200+ kilobase-pair range, apply molecular sequencing for novel clinical applications, find molecular alterations missed from matched formalin-fixed based specimens, and identify metabolites and secreted extracellular vesicles from tumor enriched cells. The key instrumental benefit of these tools is that the originating tissue source remains intact for traditional pathology evaluation, obviating the longstanding impediment to specimen access often encountered with laboratories.

POSTER

160

Precision microbiome for precision health: a strain-aware operon-resolution association platform across 10,056 human gut metagenomes

Chunyu Zhao[1,2], Anastasia Godneva[3], Eran Segal[3], Katie Pollard[2,4,5]

1. Chan Zuckerberg Biohub, San Francisco, CA, USA
2. Gladstone Institutes, Data Science and Biotechnology, San Francisco, CA, USA
3. Weizmann Institute of Science, Department of Computer Science and Applied Math, Rehovot, Israel
4. University of California, San Francisco, Biomedical Informatics Graduate Program, San Francisco, CA, USA
5. University of California, San Francisco, Department of Epidemiology and Biostatistics, San Francisco, CA, USA

Linking gut microbiome variation to host physiology is central to precision health, yet most microbiome-wide association studies (MWAS) operate at the species or strain level, where thousands of co-inherited genes collapse into a single strain-level signal, leaving the functional unit of variation unresolved. Understanding which microbial genotypes track host phenotypes, and at what level of genetic organization is increasingly important for turning microbiome measurement into clinical insight.

We developed an end-to-end MWAS platform that proceeds from raw metagenomic reads to operon-level association. The framework integrates MIDAS species-level pangenome profiling to recover accessory gene content, microSLAM relatedness-aware testing to separate strain-linked from trait-specific signals, and a strain-aware operon analysis that de novo reconstructs trait-gene operons across strains and resolves the trait-associated gene blocks within them.

We applied the MWAS platform to 10,056 participants in the Human Phenotype Project. It profiled accessory genes across 127 gut species and tested them against 80 quantitative host traits, spanning cardiometabolic, hematological, hepatic, renal, and anthropometric measures. To limit demographic confounding and depth-related detection bias, we adjusted for age, sex, BMI, medication, and sequencing depth. After correcting for strain population structure, this yielded 1,067 gene-trait associations. The strongest recurring associations did not converge on a single canonical pathway or one headline gene-trait pair. Instead, they were repeatedly found in accessory loci related to bacterial cell-surface, membrane, and envelope biology, often in genomic neighborhoods with signatures of mobility. Operon-level resolution organized these into recurring modules: surface glycan biosynthesis and utilization were associated with cardiometabolic traits, while mobilome- and prophage-associated loci were enriched among hematological traits. Much of the signal fell in otherwise unannotated “dark matter,” which recurring operon structure helped place in genomic context.

Together, these results show that host-associated microbiome variation is frequently encoded through both strain-linked genetic backgrounds and modular, often mobile, gene content. A reads-to-operon MWAS platform makes this layer measurable at cohort scale, resolving each signal at its natural level of genetic organization and turning a vast microbial search space into a tractable shortlist of candidate genes and operon modules.

POSTER

161

Predicting genetic regulatory effects in unmeasured contexts via integration of prior eQTL data with neural network models of sequence and expression

Fuchen Li[1], Ben Strober[2,3], Alexis Battle[1]

1. Johns Hopkins University, Department of Biomedical Engineering, Baltimore, MD, USA
2. Boston Children's Hospital, Computational Health Informatics Program, Boston, MA, USA
3. Harvard Medical School, Department of Pediatrics, Boston, MA, USA

Understanding how genetic variants modulate gene expression across biological contexts can reveal disease mechanisms, yet eQTL studies in many disease-relevant cellular states remain unmeasured or underpowered. It remains impractical to collect large cohorts for every combination of cell type, temporal stage, and environmental condition. To address this gap, we developed a computational framework that can predict context-specific eQTL effect sizes based on only one or a few expression samples from a target context. The framework integrates two complementary sources of information: transcriptomic state, which defines the cellular environment in which regulation occurs, and DNA sequence, which encodes the cis-regulatory code through which variants act.

Our framework first infers a low-dimensional representation of the target context from its gene expression profile. This context representation is mapped to eQTL effect sizes through a factorized model learned from large-scale eQTL resources, using a predictive framework that overcomes limitations of previous approaches that rely on fixed factors estimated only in previously measured conditions. In leave-one-tissue-out evaluation across 50 GTEx tissues, this expression-informed model significantly outperformed a baseline using eQTL effects from the closest matched tissue. To complement this context-level backbone, we use sequence-to-function neural network models to extract variant-level regulatory features from local DNA sequence, enabling prediction of effects for rare or previously unobserved variants and capturing sequence patterns such as context-specific regulatory motifs. Fine-tuning these models on existing eQTL resources, together with explicit adjustment for local linkage disequilibrium, further improves predictive performance.

We integrate expression-derived context embeddings and sequence-derived variant embeddings in a unified model that predicts genetic effects in each context, trained on a large catalog of context-specific eQTLs. We evaluate these integrative models using held-out eQTL studies to assess generalization across unseen contexts. Ultimately, this framework provides a robust approach to predict causal genetic effects in novel conditions, dramatically expanding the range of disease-relevant contexts in which genetic effects can be analyzed.

Preterm preeclampsia genome-wide association study (GWAS) of 12,793 genomes identifies a novel, ancestry-specific risk locus

POSTER

162

Kathryn J. Gray[1], Chang Cao[2,3], Sagal Dirie[2,3], Matthew Maher[2,3], Jesse Valliere[2,3], Jie Hu[2,3,4], Claudia Colmenares[5], Norma Serrano[5], Nathan Blue[6], Richa Saxena[2,3]

1. University of Washington, Department of Obstetrics & Gynecology, Seattle, WA, USA
2. Massachusetts General Hospital, Center for Genomic Medicine, Boston, MA, USA
3. Broad Institute of MIT and Harvard, Cambridge, MA, USA
4. Harvard T.H. Chan School of Public Health, Boston, MA, USA
5. Fundación Cardiovascular de Colombia, Bucaramanga, Santander, Colombia
6. University of Utah Health, Salt Lake City, UT, USA

Preeclampsia (PE) is a common, severe pregnancy disorder characterized by new-onset hypertension and proteinuria after 20 weeks. PE confers an increased lifelong risk of cardiovascular disease (CVD) and is highly heritable (55%; 30% maternal, 25% fetal). Published PE GWAS lack ancestral diversity and have not considered PE subtypes, including PE delivered <37 weeks (preterm PE; a severe subtype with increased morbidity). We utilized TOPMed whole genome sequencing (WGS) data from the Boston-Colombia Collaborative for Adverse Pregnancy Outcomes (BCC-PREG) cohort and nuMoM2b-Heart Health Study (11,520 mothers (1,651 preterm PE), 1,273 offspring (275 preterm PE) (47% AMR, 33% EUR, 13% AFR, 7% other)), for a preterm PE GWAS. Genome-wide association testing was performed using SAIGE with co-variate adjustment. The preterm PE maternal GWAS identified one genome-wide significant SNP on chromosome 11 near *JRKL* (OR=1.28, [95% CI 1.18-1.39]; p=7.40e-9) with the largest effect in the admixed American (AMR) and African populations in sub-ancestry analyses (AMR: OR=1.21 [1.08-1.35], p=7.96e-4; EUR: OR=1.20 [0.99-1.45], p=6.27e-2; AFR: OR=1.54 [1.24-1.91], p=1.01e-4). The leading maternal SNP was nominally significant in the preterm PE fetal GWAS (OR=1.52 [1.21-1.92], p=3.51e-4). Conditional genetic effect analysis revealed independent effects of both the maternal and fetal genome at this locus, suggesting both influence risk. JRKL is a nuclear regulatory protein expressed in most human tissues, including placental trophoblasts and endothelial cells, but its function has not yet been delineated. As PE confers a future risk of cardiometabolic disease in affected patients, polygenic risk scores (PRS) for 20 cardiometabolic traits were generated using PRS-CSx and tested for association with preterm PE. PRS for stroke, coronary artery disease, BMI, type 2 diabetes, and blood pressure traits were significantly associated with preterm PE—the stroke association was the strongest (OR=1.56 [1.30–1.90], p=7.45e-6). In this multi-ancestry preterm PE GWAS of maternal and fetal genomes, we identified a novel PE genetic risk locus with ancestry-specific effects. We also found an unexpectedly strong association between genetic predisposition to stroke and preterm PE, supporting shared underlying biological mechanisms. This work highlights the importance of including diverse populations and clinical subtyping in PE genetic studies and delineates pathways for future investigations.

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

POSTER

163

Neil Slaven[1], Sarah Morton[2], Michal Kosicki[1], Laura E Cook[1], Rashedul Islam[1], Momoe Kato[1], Ingrid Plajzer-Frick[1], Christine Seidman[2], Jonathan Seidman[2], Len Pennacchio[1], Axel Visel[1]

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

Congenital heart disease (CHD) is the most common birth defect worldwide, affecting millions of individuals in the US alone. Despite significant research efforts, coding variants are estimated to account for only 30% of cases. Seminal studies have demonstrated that non-coding variants in regulatory elements, such as transcriptional enhancers, can contribute to the development of congenital diseases. It is predicted that some unresolved cases may be explained by non-coding variation; however, the overall contribution of heart enhancer variants to CHD remains unclear. To address this knowledge gap, we aimed to prioritize non-coding variants from a unique CHD patient cohort and characterize their functional impact *in vivo*. Whole-genome sequencing was performed by the Pediatric Cardiac Genomics Consortium (PCGC) on a cohort of 1,831 trios with severe cardiac defects and no evidence of pathogenic coding mutations, identifying >75,000 *de novo* and >1 million rare variants in the non-coding genome. Intersection with a predicted heart enhancer catalogue identified ~1,000 *de novo* and ~220,000 rare candidate variants. Variants were prioritized using multiple lines of evidence: H3K27Ac enrichment, open chromatin, evolutionary conservation, and predicted disruption of cardiac transcription factor binding sites. Functional testing of wildtype and variant enhancer sequences was performed *in vivo* using a reporter knock-in mouse assay (enSERT). To date, we have identified 20 novel developmental heart enhancers, of which 7 (35%) displayed altered activity upon variant introduction. Application of a deep learning model (ChromBPNet) revealed strong concordance between predicted variant effects and observed *in vivo* activity. Together, these data demonstrate that non-coding variants frequently (35%) disrupt heart enhancer activity and that deep learning models can accurately predict these effects.

Putting two exome/genome sequencing reanalysis programs to work in an orphan disease cohort

POSTER

164

Elly Brokamp[1]

1. Boston Children's Hospital, Boston, MA, USA

Iterative analysis of non-diagnostic exome and genome sequencing (ES/GS) testing as genomic knowledge evolves is an ongoing challenge for clinicians and researchers. Scalable ES/GS reanalysis approaches, including those that incorporate Artificial Intelligence (AI), may relieve manual effort and time required for reanalysis. We investigate the performance of two reanalysis programs that prioritize variants for manual review in a large rare-disease cohort. One tool, Talos, emphasizes incorporating updated knowledge sources, such as ClinVar and PanelApp, for continual variant-prioritization as these evidence sources update. The other tool, Variant Scoring Neural Network (VS-NN), is an AI-assisted approach that emphasizes utilizing patient phenotypic data to prioritize appropriate disease-associated genes of interest from OMIM.

ES/GS proband/ trio data were evaluated using Talos (2,913 probands) and VS-NN (388 probands) with a list of candidate variants (CVs) returned for manual review. Probands were children and adults with suspected monogenic disease whose ES/GS data had been previously analyzed either by the research team or a clinical laboratory. Trained variant curators reviewed all CVs, determining if they are non-diagnostic, a candidate diagnosis needing further workup, or diagnostic. We evaluated the CV burden for manual review and differences between variants prioritized by Talos and VS-NN.

Talos flagged at least one CV for 1,313 probands with ES and 333 probands with GS averaging 2.1 (ES) and 2.5 (GS) CVs per proband. VS-NN flagged at least one CV for 18 probands with ES and 19 probands with GS averaging 2.9 (ES) and 3.5 (GS) CVs per proband.

Through GS reanalysis, Talos has assisted in making at least eight new diagnoses and eighteen potential diagnoses needing further investigation. VS-NN has assisted in 1 potential new diagnosis, which was not prioritized by Talos.

Among 47 probands with a previously known diagnosis via GS, Talos flagged the diagnostic variant in 18 (38.3%). Among 30 probands with a previously known diagnosis via GS, VS-NN flagged 4 (13.3%). The known diagnoses not flagged by Talos/VS-NN were mostly structural variants, complex rearrangements, repetitive elements, variants needing phasing data, or novel gene-disease correlation.

In a large rare-disease cohort, reanalysis tools facilitated identification of new and potential diagnoses at scale.

POSTER

165

scVIGR: interpretable single-cell variational inference of gene regulation for perturbation prediction

Derek E. Bowman[1], **Sudin Bhattacharya**[1]

1. Michigan State University, East Lansing, MI, USA

Introduction: Precision medicine requires predicting how an individual's biology will respond to treatment or other perturbations and identifying the regulatory mechanisms governing those responses. Variational autoencoders (VAEs) can predict single-cell transcriptional responses by learning latent representations that support vector arithmetic across conditions and cell types, but these latent spaces are difficult to interpret biologically. We therefore integrated deep generative modeling with gene regulatory network inference to enable interpretable perturbation prediction.

Methods: scVIGR retains scVIDR's variational encoder and latent perturbation arithmetic but replaces its generic decoder with a transcription factor (TF)-activity bottleneck and gene regulatory network-constrained decoder. Cell-type-specific perturbation vectors are applied to held-out control cells, and predicted latent states are decoded through non-negative TF activities. CollecTRI-covered genes are predicted through a masked, sign-aware TF–gene layer, while uncovered genes are modeled through an L1-regularized discovery layer. We evaluated predictions for central hepatocytes exposed to 30 μg/kg TCDD and CD14+ monocytes stimulated with IFN-β. Interpretability was assessed using degree-matched randomized-GRN controls and ReMap2022 ChIP-seq support.

Results: scVIGR reconstructed both perturbation responses while recovering coherent regulatory programs. In TCDD-treated central hepatocytes, it inferred increased Ahrr activity (+7.2) and decreased Hnf4g (−7.0), Hnf1a (−6.0), and Srebf1 (−5.0), consistent with AhR activation and suppression of hepatocyte-identity and lipid-regulatory programs. In IFN-β–stimulated CD14+ monocytes, STAT1 showed the largest increase (+13.4), accompanied by STAT2, IRF9, and multiple IRF-family regulators, recovering the canonical type-I interferon response. Discovery edges linked these regulators to established interferon-stimulated genes, including *IFI6, ISG15, IFIT1, IFIT3, RSAD2, CXCL10,* and *CXCL11*. Randomized-GRN models reconstructed expression similarly but failed to recover expected regulators and showed lower ReMap corroboration, indicating that interpretability depended on the biological content of the GRN prior.

Conclusion: scVIGR adds mechanistic interpretability to deep generative perturbation modeling while retaining strong predictive performance, enabling analysis of regulatory shifts underlying cellular responses.

Sequential decision making for pedigree-based genetic testing

POSTER

166

Kanix Wang[1], Daniel Adelman[2]

1. University of Cincinnati, Carl H. Lindner College of Business, Department of Operations, Business Analytics, and Information Systems, Cincinnati, OH, USA
2. The University of Chicago Booth School of Business, Operations Management, Chicago, IL, USA

Multigene panel testing can reveal hereditary risk across a family, but it creates a sequencing problem. Testing one relative changes what can be learned from others. The number of possible testing histories grows rapidly with pedigree size and number of genes. We asked whether pedigree wide planning changes whom to test and whether it can ultimately scale to clinical gene panels.

We modeled Mendelian inheritance for panels containing multiple genes and used sequential optimization to choose after each result whether to stop or test another relative. The objective balanced information about familial carrier status against testing costs. It can accommodate measures of clinical benefit or QALYs. We compared the model optimal strategy with a one step strategy across multiple simulated scenarios spanning pedigrees, genetic assumptions, information values, and testing costs. In 102 settings containing multiple genes with exact solutions, we evaluated an approximation algorithm intended to reduce exhaustive calculation. To extend further, we developed an AI approach via reinforcement learning based on a graph neural network that represents family structure and gene level information across variable pedigree and panel sizes.

The one step strategy lost a meaningful amount of model value in majority of scenarios. In some, it selected a different first relative than the model optimal strategy. In another, it selected the same first relative but made inferior later decisions. Across the 102 exact benchmarks, the approximation algorithm reduced average decision loss by 20.8% and reduced error in the upper bound used to verify solution quality by 69.4% compared with naïve approaches.

These results show that the value of a genetic test depends not only on the person tested now, but also on how that result changes later choices for the family. The study provides a proof of principle for pedigree-wide planning in genetic testing.

POSTER

168

Streamlined workflow generates small RNA and RNA-seq libraries from the same sample for deeper insights into tumor and normal matched tissues

Alexis Kokaras[1], Heather M. Raimer Young[1], Gautam Naishadham[1], Bradley W. Langhorst[1], Louise Williams[1]

1. New England Biolabs, Ipswich, MA, USA

Small noncoding RNAs (sncRNAs) play fundamental roles in many biological processes including cancer development and drug resistance. Increasing evidence shows that sncRNA expression signatures can be used to discriminate between normal and cancer tissues, highlighting their potential as diagnostic and prognostic biomarkers of the disease. Differentially expressed sncRNAs can be investigated using small RNA next generation sequencing methods and can be used to infer downstream effects on the transcriptome. However, performing RNA-seq from the same samples is necessary to confidently determine expression of any given affected transcript and understand the overall transcription profile. Making both small RNA and standard RNA-seq libraries can be challenging when sample material is limiting. Workflows that enable multiple data types to be generated from a single sample are needed for comprehensive understanding of disease.

Here we demonstrate the use of the NEB Monarch RNA size fractionation workflow to allow a single sample to be used for both small RNA and ribosomal depleted RNA-seq library preps. The Monarch RNA cleanup columns were used to enrich small RNA (< 200 nucleotides) and recover large transcripts (> 200 nucleotides) from 6 different human tissues (brain, testis, placenta, bladder, ovary, and esophagus) as well as 5 sets of matched normal and tumor samples (breast, stomach, rectum, colon, and liver).

Small RNA libraries were prepared using either total RNA or enriched small RNA samples (from the Monarch RNA column) with the NEBNext® Low-bias Small RNA Library Prep Kit. Ribosomal depleted RNA-seq libraries were prepared from the large transcript fractions using the NEBNext UltraExpress® RNA Library Prep Kit downstream of NEBNext rRNA Depletion Kit v2. We correlated expression of individual sncRNAs between total RNA and enriched small RNA libraries across all six normal tissues. In matched tumor and normal samples, we evaluated the relationships between miRNAs and their targets within the differentially expressed transcripts from both small RNA and transcriptome datasets. Overall, generation of small RNA and RNA-seq libraries from the same starting material permits a better understanding of the biological regulation of transcriptomes in individual samples.

Sunshine Genetics: enabling scalable genomic newborn screening in Florida through Nest Genomics, a digital genomics platform

POSTER

169

Saanchi V. Shah[1], Emily Reeves[2], Dusty Stern[2], Susanne Crowe[3], Patricia Ryland[3], Sharon Liebrich[1], Pradeep G. Bhide[1], David H. Ledbetter[1], Moran Snir[4]

1. Florida State University College of Medicine, Florida Institute for Pediatric Rare Diseases, Tallahassee, FL, USA
2. Florida Department of Health, Newborn Screening Follow-Up Program, Tallahassee, FL, USA
3. Florida Department of Health, Bureau of Public Health Laboratory, Jacksonville, FL, USA
4. Nest Genomics, New York, NY, USA

Background: Genomic newborn screening (gNBS) programs are rapidly expanding globally, enabled by next-generation sequencing technologies that identify hundreds of conditions beyond traditional newborn screening (tNBS). Yet, integration into tNBS programs remains limited due to challenges in workflow harmonization, data capture, and gaps between early detection and access to clinical specialists. The Sunshine Genetics (SG) Act, effective July 2025, established a public-private partnership to implement a statewide gNBS program in Florida (FL). SG is designed as a digitally enabled model to address workforce and infrastructure limitations.

Methods: SG is a prospective study enrolling up to 100,000 FL newborns aged ≤30 days. A digital genomics platform, Nest Genomics (Nest), supports 3 phases: (1) Pre-enrollment: pre-and-postnatal education and asynchronous consent; (2) Post-enrollment: linkage to dried blood spot cards via FL Public Health Lab workflows and Application Programming Interface (API)-based genetic test ordering with sequencing partners; (3) Results disclosure and follow-up using telegenetics supported by artificial intelligence (AI)-enabled care-plan templates and SG ACTion Sheets for providers and families. Nest provides alerts for discordant tNBS and gNBS findings, supports urgency-based care coordination. An AI-powered question-and-answer feature provides immediate responses to inquiries, with the option to connect with the centralized hub at the FL Institute for Pediatric Rare Diseases (IPRD).

Results: SG integrates end-to-end gNBS workflows within a single digital platform designed for scalable statewide implementation and follow-up. SG screens up to 918 pediatric-onset conditions using a two-tiered panel based on whole genome sequencing, with reporting of pathogenic and likely pathogenic sequence and copy number variants. Participants access data through Nest, “Florida Passport”, for ≥7 years enabling genomic data reuse. The Nest-Amazon Web Services instance interfaces with the IPRD data warehouse, allowing cloud storage of demographic, clinical, and follow-up data facilitating federated computational analyses.

Conclusion: SG will evaluate the usability and effectiveness of Nest for FL-wide gNBS implementation. Health economic analyses will assess whether digital and telegenetics workflows reduce workforce constraints and generate long-term cost savings, informing implementation of gNBS as a population-scale public health program in FL.

Systematic gaps in reporting variants of uncertain significance (VUS) and their reclassifications

POSTER

170

Pankhuri Gupta[1,2], Andrew Folta[2], Adriana Sedeno Cortes[2], Eric Kao[2], Abbye E. McEwen[1,3], Martha Horike-Pyne[1], Jeremy Stone[1], Douglas M. Fowler[1,4], Lea M. Starita[1,4], Andrew B. Stergachis[1,2,4]

1. University of Washington, Department of Genome Sciences, Seattle, WA, USA
2. University of Washington, Division of Medical Genetics, Seattle, WA, USA
3. University of Washington, Laboratory Medicine and Pathology, Seattle, WA, USA
4. Brotman Baty Institute for Precision Medicine, Seattle, WA, USA

Variants of uncertain significance (VUS) are frequently encountered during clinical genetic testing, yet the clinical burden they impose on patients and the dissemination of VUS reclassifications remain largely unresolved. To investigate these, we developed the Brotman Baty Institute Clinical Variant Database (BBI-CVD), an electronic health record (EHR)-linked database of clinical germline genetic variant information from adult patients seen at the University of Washington and Fred Hutchinson Cancer Center.

Through retrospective chart review, we characterized the distribution of VUS burden in 5,158 participants across testing indications and self-reported race. Variant classifications in the BBI-CVD were also systematically compared to ClinVar classifications to identify discordant classifications and identify gaps in dissemination of these reclassifications. All discrepancies were manually reviewed to ensure that the variant classifications applied to the same disease entity for which the patient was tested. Additionally, participants (n=20) who had previously received variant reclassifications were interviewed to explore perspectives on receiving reclassification results. Template analysis on the interview transcripts using inductive and deductive codes was utilized to identify major themes.

The number of reported VUS relative to P/LP variants varied by over 14-fold depending on the primary indication for genetic testing and by 3-fold depending on self-reported race, indicating that VUS burden is unequally distributed across patient populations. Comparison with ClinVar revealed that 3.4% (176 variants) carried a substantively different classification relative to the EHR, including 26 instances in which the testing laboratory updated ClinVar but did not notify the patient or the provider. Interviews with participants demonstrated that they viewed variant reclassifications as evidence of personalized medicine, and at the same time highlighted dissatisfaction with the method and timing of result disclosure.

The findings underscored that the clinical burden of VUS in adult medical genetics is unequally distributed, with systematic biases determining which patients are more likely to receive VUS results. The findings also reveal a critical communication gap in the dissemination of variant reclassifications. Systematic methods are needed to ensure smooth, timely delivery of results to patients and providers for clinical management and to reduce patient burden.

The newborn profile: a clinical framework enabling genome-informed care at birth

POSTER

171

Whitney S. Thompson[1], Laura M. Amendola[2], Cheyenne Camp[2], Meredith Wright[2], Amie E. Jones[1], Monica Giovanni[2]

1. Mayo Clinic Children's, Rochester, MN, USA
2. Inflection Medicine, Boston, MA, USA

Newborn screening (NBS) is a widely accepted and successful population health program. Over the last decade, research initiatives have evaluated genomic sequencing in NBS to expand condition detection and accelerate diagnosis. Building on these efforts, the Newborn Profile (NBP) was developed to translate genomic newborn screening into clinical practice, enabling identification of actionable childhood-onset conditions within a clinical care framework aligned with established NBS principles.

A multidisciplinary team with expertise in clinical and biochemical genetics, neonatology, complex care pediatrics, laboratory medicine, implementation science, and bioethics developed a structured methodology for NBP condition selection. Criteria included medical actionability, gene-disease validity, diagnostic accuracy, feasibility of confirmation, and alignment with existing NBS paradigms. Application of these criteria identified approximately 450 disorders with opportunities for intervention during infancy or early childhood. Use of a blended short- and long-read genome backbone for the NBP enables inclusion of some conditions historically excluded from genomic NBS due to technical limitations.

Integration of molecular findings with confirmatory testing and seamless transition to phenotype-driven diagnostic genome sequencing without additional sample collection distinguish the NBP from research-based genomic NBS programs. All conditions were evaluated for the availability and performance of orthogonal confirmation methods and incorporated into a tiered disclosure workflow that enables confirmation before result return when appropriate, while fast-tracking urgent conditions or those without confirmatory testing options. Embedded clinical decision support further facilitates immediate clinical action.

A governance framework was also developed for ongoing condition-list maintenance. Periodic evidence review incorporates emerging genomic NBS programs, gene-disease validity, laboratory advances, and new or repurposed therapies. Conditions may also be nominated by clinicians, researchers, and patient advocacy groups. Inclusion is guided by predefined criteria and reassessed at a regular cadence to maintain clinical relevance and technical feasibility.

Together, these efforts establish a scalable, adaptable framework for translating advances in genomic medicine into clinical newborn screening. Implementation at scale will require a cost-effective model that can be seamlessly integrated into frontline clinical care.

POSTER

172

Trio-aware long-read genome sequencing with multi-omic integration resolves previously unsolved rare disease cases

Daniel Calame[1], Nikhita Gogate[1], Medhat Mahmoud[1], Moez Dawood[1], Shannon Dugan-Perez[1], Heer Mehta[1], Yiming Zhu[1], Chloe Munderloh[1], Shaghyegh Beheshti[1], Yilei Fu[1], Ryan Cheng[1], Ryan German[1], Christopher Grochowski[1], Donna Muzny[1], James Lupski[1], Richard Gibbs[1], Fritz Sedlazeck[1], Jennifer Posey[1]

1. Baylor College of Medicine, Houston, TX, USA

Despite the widespread uptake of short-read exome/genome sequencing into clinical diagnostics, a substantial proportion of individuals with suspected Mendelian disorders remain undiagnosed. Long-read genome sequencing (LR-GS) improves resolution of structural variants, repeat expansions, and repetitive, complex regions while simultaneously providing genome-wide methylation profiling, yet its diagnostic contribution in clinically unsolved cohorts remains incompletely defined. We applied LR-GS to 923 individuals from 285 families, including 167 trios, using PacBio and Oxford Nanopore platforms. Our analytical framework incorporated >600 genomic and functional annotations across single-nucleotide variants, indels, structural variants, repeat expansions, and methylation data, integrated with short- and long-read transcriptomics and untargeted metabolomics in a subset of cases.

We established molecular diagnoses in 56/285 families (19.6% yield) in this unsolved cohort. Critically, 13 diagnoses (23% of molecular findings) were accessible only through LR-GS resolving complex variants undetected by short-read sequencing. These LR-GS-dependent diagnoses comprised structural variants, repeat expansions, and deep-intronic variants; the full set included 35 coding SNVs, 6 splice-altering or deep-intronic SNVs, 1 indel/frameshift, 12 structural variants, and 2 tandem repeat expansions. Highlights included 1) an intergenic inversion of the HOXC8 and PDE1B locus segregating with HTT negative chorea, producing a chimeric PDE1B transcript in patient-derived induced neurons predicted to truncate the N-terminal regulatory region; 2) a complex WWOX rearrangement – an intragenic inversion flanked by deletions – in trans with a likely pathogenic missense variant in a child with epileptic encephalopathy; and 3) bi-allelic NAXE promoter GGGCC repeat expansions causing a treatable mitochondrial encephalopathy. Additional LR-GS-dependent diagnoses included a TRIP12 tandem duplication and compound heterozygous TSPEAR variants comprising a missense allele and a two-exon deletion. Trio-based segregation and multi-omic integration were essential for interpretation, in one case delivering a treatable diagnosis in a proband and pauci-symptomatic sibling that standard testing had missed. Trio-aware LR-GS with multi-omic integration is a logical next-tier test after non-diagnostic short-read sequencing, adding a clinically meaningful diagnosis in 4.6% of unsolved families.

Ultra-rapid sequencing needs faster humans: prenatal consent for NICU genome sequencing

POSTER

173

Rebecca Reimers[1,2], Brandan Schultz[1], Hannah Bressner[2], Corrine Blucher[1], Kristen Wigby[1,3], Katarzyna Ellsworth[1]

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

Genome sequencing (GS) in neonatal and cardiac intensive care units (NICU/CICUs) has transformed care for neonates with congenital anomalies and been shown to improve cost effectiveness of care in prior studies. The speed of sequencing has continued to accelerate but often the human-side of technology implementation lags.

Rady Children's Hospital, San Diego (RCHSD) implemented a quality improvement process of identifying pregnant persons who planned to deliver a neonate with single ventricle physiology (e.g., hypoplastic left heart). Starting in August of 2025, families were approached for consent for neonatal GS prenatally after consultation with the fetal cardiology team. This process was supervised by a Maternal-Fetal Medicine/ Clinical Geneticist and Fetal Cardiologist and coordinated by a neonatal cardiac nurse coordinator and NICU genetic counselor. The consent conversation was documented in the pregnant person's chart and blood or buccal samples were obtained from the parents prenatally, DNA extracted, and then stored. GS after birth followed standard clinical lab procedures.

15 patients were eligible in the time frame, of those 11 had prenatal consent for a completed postnatal genome. Of those, 10 patients had GS ordered by clinicians on admission to the hospital which resulted at an average of 2.4 days (st.dev 1.6) from birth to sample collection. One patient had nondiagnostic exome sequencing and microarray prenatally and thus clinicians did not initially order GS. However, GS subsequently ordered at10 days of life was nondiagnostic. Three patients did not have prenatal consent, two were consented postnatally with average time from birth to collection of 3.5 days. One patient was never consented and the neonate died at four days of life without any genetic testing. In the same time frame, neonates admitted to the CTICU following birth were ordered for ultrarapid GS an average of 9 days after birth (st.dev 8) in those who did not have prenatal consent (p=0.03).

Opportunities for improvement in rapid diagnosis of critically ill neonates continue to advance- both on the human and technical sides of care provision. We identified and trialed a process of prenatal consent for neonatal genome which reduced the time to GS collection by 5.5 days.

Unraveling the mechanisms of RHOBTB2-related developmental and epileptic encephalopathy

POSTER

174

Athena Olszewski[1,3], Davneet Kaur[1], Jacob A. Steele[3,4], Soham Sengupta[1], Shondra M. Pruett-Miller[2,3,4], Heather C. Mefford[1,2]

1. St. Jude Children's Research Hospital, Genomic and Translational Neuroscience, Memphis, TN, USA
2. St. Jude Graduate School of Biomedical Sciences, St. Jude Children's Research Hospital; Memphis, TN, USA
3. St. Jude Children's Research Hospital, Department of Cell and Molecular Biology, Memphis TN, USA
4. St. Jude Children's Research Hospital, Center for Advanced Genome Engineering, Memphis TN, USA

Heterozygous missense variants in the gene *RHOBTB2* cause a developmental and epileptic encephalopathy (RHOBTB2-DEE) characterized by epilepsy, intellectual disability, and movement disorders. Three recurrent variants, R483H, R511Q, and R511W, account for ~70% of RHOBTB2-DEE cases. Individuals with heterozygous *RHOBTB2* loss-of-function (LOF) variants are healthy, whereas those with biallelic LOF variants have neurological presentations that are generally milder than RHOBTB2-DEE. Based on the genotype-phenotype relationship, we hypothesize that DEE variants in *RHOBTB2* act via a gain-of-function or dominant-negative mechanism. Understanding the effect of RHOBTB2-DEE variants is important for designing disease models for preclinical testing of therapeutics, identifying novel drug targets and possible disease biomarkers, and improving variant impact prediction. Our investigation aims to determine the molecular and cellular consequences of RHOBTB2-DEE variants and evaluate the therapeutic potential of antisense oligonucleotides (ASOs) for RHOBTB2-DEE. We created a set of isogenic SF126 glioblastoma cell lines with engineered *RHOBTB2* variants (WT/WT, WT/-, WT/R483H, WT/R511Q, WT/R511W). In an assay to capture ER-to-Golgi transport dynamics, RHOBTB2-DEE cells exhibited increased localization of a tracker protein at the Golgi. To identify more cellular processes affected by RHOBTB2-DEE, we performed RNA-seq. Comparison between DEE and non-DEE lines revealed downregulation of genes involved in the regulation of neuronal projection development, extracellular matrix organization, integrin binding, and axonogenesis. RHOBTB2-DEE is a good candidate for allele-specific knockdown using an ASO since individuals with one LOF allele do not have DEE. We designed gapmer ASOs targeting recurrent RHOBTB2-DEE variants in an effort to design a therapeutic that could be used across multiple patients. We identified several candidate ASOs that cause knockdown of *RHOBTB2* in SF126 models, however follow-up testing of cells treated with the lead ASO targeting the R483H variant revealed that *RHOBTB2* knockdown was, surprisingly, not allele-specific. Additionally, genes differentially expressed between RHOBTB2-DEE and non-DEE cell lines did not show a shift toward non-DEE levels after ASO treatment. Our research suggests that intracellular transport and neuronal morphology may be *in vitro* phenotypes of RHOBTB2-DEE that may be used as biomarkers for screening ASOs in preclinical studies of additional ASOs targeted to RHOBTB2-DEE alleles.

Veterinary Undiagnosed Disease Initiative: rare disease translation between humans and animals

POSTER

175

Anne S. Maguire[1], My H. Hoang[2], Emily C. Graff[1], Douglas R. Martin[1]

1. Auburn University College of Veterinary Medicine, Scott-Ritchey Research Center, Auburn, AL, USA
2. Washington University School of Medicine, St. Louis, MO, USA

The purpose of the Veterinary Undiagnosed Disease Initiative (VUDI) is to improve health outcomes for people and animals with rare disease. Recent advances in therapeutic strategies for rare disease would benefit from validation in animal models that bridge the gap between rodents and humans. Cats and dogs have a long history as intermediate models, providing more accurate translation of size, disease progression, and response to treatment. However, while ~7000 causative variants of rare disease have been identified in humans, only several hundred have been identified in animals. The goal of the VUDI is to build the infrastructure to close this gap, thereby offering 1) genetic diagnoses and potential therapies for animals with rare disease, and 2) a pool of affected animals as candidates for therapeutic development.

Specifically, the VUDI will diagnose new cases of rare disease in companion animals, then maintain a database of the causative variants, with the ultimate goal of connecting owners of affected animals with partners that are developing appropriate therapeutics. Owners receive hope in the form of a promising new therapeutic, while pharmaceutical companies receive valuable information about how their therapies work in a large animal model. This initiative builds upon Scott-Ritchey Research Center's decades-long history of working closely with industry to develop gene therapy for lysosomal storage disorders (LSDs).

To determine the feasibility and impact of this initiative, we are developing pilot projects in dogs with epilepsy and cats with LSDs. We are identifying new, monogenic, causative variants for 10 dogs diagnosed with idiopathic epilepsy. We use PacBio long-read sequencing at >20x coverage and a bioinformatics pipeline modeled after established practices in the Undiagnosed Diseases Network for human patients. For cats diagnosed with LSDs through traditional metabolic tests, we use short-read sequencing at >30x coverage to characterize novel variants.

By connecting all stakeholders who are passionate about genomics and translational medicine, the VUDI will improve the diagnosis, treatment, and quality of life for people and animals with rare disease.

POSTER

176

Ancestry shapes the epigenetic, regulatory, and immune architecture of Alzheimer's disease: findings from the MAGENTA study

William S. Bush[1], Matthew H. Liu[1], Makaela Mews[1], Yousef Mustafa[1], Tianjie Gu[2], Lissette Gomez[2], Sebastián Cruz-González[3], John A. Capra[3], Margaret A. Pericak-Vance[2], Jonathan L. Haines[1], Anthony Griswold[2]

1. Case Western Reserve University, Population and Quantitative Health Sciences, Cleveland, OH, USA
2. University of Miami, Miller School of Medicine, John P. Hussman Institute for Human Genomics, Miami, FL, USA
3. University of California, San Francisco, Department of Epidemiology and Biostatistics, San Francisco, CA, USA

Alzheimer's disease (AD) is more prevalent among African American and Hispanic populations than among non-Hispanic Whites, yet multi-omics studies of AD risk remain overwhelmingly European in composition. The Multi-Ancestry Genomics, Epigenomics, and Transcriptomics of Alzheimer's (MAGENTA) study addresses this gap, generating whole-blood genotyping, DNA methylation, and RNA sequencing data from African American, Caribbean Hispanic, Peruvian, and non-Hispanic White AD cases and cognitively normal controls ascertained under harmonized protocols. We identify substantial differences in peripheral immune -omics measures by ancestry and AD status.

First-, second-, and third-generation DNA methylation clocks predicted chronological age less accurately in those with African genetic ancestry (Horvath r = 0.51 in African Americans and 0.45 in Puerto Ricans versus 0.72 in Whites). This disparity replicates in ~2,500 external individuals, and most clocks failed to detect age acceleration in admixed AD cases. Nearly a quarter of Horvath clock CpGs were differentially methylated between African and European ancestry individuals, and clock CpG methylation QTLs were markedly more frequent on African ancestry backgrounds.

When compared to similar multi-ancestry studies, we identify 83,633 novel eQTL associations, including between 1,436 and 4,733 AD-responsive cis-eQTLs (ADR-eQTLs) per population. Just 3.3% of ADR-eGenes were shared across populations, compared with 24% of baseline eGenes, highlighting the population-specific nature of immune response to AD. ADR-eQTLs localized outside gene bodies and were enriched for chromatin states, histone modifications, and promoter-capture Hi-C interactions.

Further supporting these immunological differences, deconvolution of whole-blood transcriptomes indicated AD-associated immune remodeling. In general, lymphoid and monocyte lineages were depleted while neutrophils expanded. Immune cell composition varied by AD status in breadth and direction across groups, with the broadest shifts in African American and non-Hispanic White participants, alongside shared interferon and heme metabolism programs and population-specific pathway signatures. We also note marked differences in the presence of clonal hematopoiesis driver mutations in this aged population, with STAG2 mutations more frequent than the commonly reported DNMT3A and TET2 (76, 60, and 23 participants, respectively). Together, these findings demonstrate that genetic ancestry shapes the epigenetic, regulatory, and immune signatures of AD, and that multi-omic signatures derived from European-ancestry data risk misclassifying the populations most affected.

High-resolution exposomic profiling of matched maternal serum and placental tissue in adverse pregnancy outcomes

POSTER

177

Tom Cohen[1], Ashima Mehta[1], Caitlin Kelly[1], Adam Richardson[1], Avery August[3], Katie Fisch[3], Ethan Stancliffe[1], Gary Patti[1,2]

1. Panome Bio, Inc., St. Louis, MO, USA
2. Washington University in St. Louis, St Louis, MO, USA
3. University of California, San Diego, San Diego, CA, USA

Background: Precision health has been built largely on the genome, yet genomic risk alone does not explain adverse pregnancy outcomes such as preeclampsia, maternal vascular malperfusion, fetal growth restriction, and preterm birth. The placenta governs the maternal–fetal interface and is highly vulnerable to environmental perturbation, yet its chemical environment remains largely unmeasured, as targeted assays capture only a narrow subset of known compounds.

Methods: We applied a dual LC/MS exposomics workflow to matched maternal serum and placental villous tissue from 50 pregnancies spanning normotensive and adverse outcome groups. Targeted Exposomics provided absolute quantification of 235 validated environmental chemicals at pg/mL sensitivity. An untargeted Discovery Exposomics workflow screened each sample against a curated database of more than 32,000 chemical exposures using stringent spectral matching and retention time prediction, enabling simultaneous xenobiotic and metabolite profiling.

Results: Discovery Exposomics detected 3,078 compounds in serum (165 chemical exposures) and 2,406 compounds in placenta (136 exposures), with 17 exposures shared across matrices. Targeted quantification recovered 30 exposures in serum and up to 36 in placenta, 15 of which were detected in both; roughly half of quantified exposures were present in both compartments. Multiple PFAS species, including PFOS, PFHxS, and PFBS, were quantified in both matrices with serum–placenta correlation for select compounds, supporting maternal circulation as a minimally invasive proxy for placental chemical burden. Detected chemicals spanned pesticides, PFAS, phthalates, plasticizers, industrial solvents, and combustion byproducts; several food and consumer product additives were detected only by untargeted profiling.

Conclusions: These data establish the placenta as a measurable archive of environmental exposure and demonstrate that high-resolution exposomics can be layered onto genomic, transcriptomic, and proteomic data to build a more complete picture of pregnancy risk. Differences between normotensive and adverse outcome groups are hypothesis generating and support exposure measurement as a tractable component of precision health.

Integrating genomic data for predicting pathology in adolescent idiopathic scoliosis

POSTER

178

Awtum M. Brashear[1], Anxhela G. Gustafson[1], Andrew Quitadamo[1], Colin Naughton[2], Nicole Nelligan[1], Natalie Groth[1], King Jordan[2], Cathleen L. Raggio[3], Brian Ciruna[4], Philip Giampietro[5], Steven Hwang[6], Kamran Shazand[1]

1. Shriners Children's Genomics Institute, Tampa, FL, USA
2. Georgia Institute of Technology, Atlanta, GA, USA
3. Hospital for Special Surgery, Uniondale, NY, USA
4. SickKids Hospital, Toronto, Ontario, Canada
5. Shriners Children's Hospital, Chicago, IL, USA
6. Shriners Children's Hospital, Philadelphia, PA, USA

Adolescent idiopathic scoliosis (AIS) is the presence of an abnormal spinal curvature of ≥ 10º in the absence of an underlying vertebral malformation or syndromic cause. Bracing is required in ~10% of cases, and ~25% of these patients will fail bracing. In some severe cases, surgical intervention is required. Patients who undergo surgery while curves are smaller and more flexible tend to have better clinical outcomes, however surgery includes substantial risk. Therefore, there is a critical need for predictive tools to help clinicians decide which patients are best suited for bracing or surgical intervention and when. To date, there is limited understanding of how genetics may inform such treatment decisions, despite mounting evidence that AIS is genetically heterogeneous. Our goal is to identify biomarkers that can improve diagnosis and guide personalized treatment approaches for AIS patients.

We performed Whole Genome Sequencing (WGS) on a cohort of 485 patients with Infantile, Juvenile or Adolescent Idiopathic Scoliosis. 431 probands (89%) had at least one family member with WGS sequencing and we retrospectively collected relevant clinical data across multiple timepoints. Family-level pedigree analysis and variant prioritization resulted in a genetic network significantly enriched for collagen trimerization and developmental processes. A cohort-level collapsing burden analysis at the pathway level similarly resulted in significant enrichment for pathways related to collagen chain trimerization, elastin fibers and mTOR signaling. Several candidates produced functional phenotypes in zebrafish. These results together suggest a connective tissue component as well as developmental timing. Our cohort is enriched for severe cases that require bracing or surgery (median Cobb angle of 38º), allowing our continued investigation into variants and pathways that influence severity and progression. We plan to integrate results from our analyses into a multi-modal model for predicting spine treatment outcomes that provide the best possible outcomes for our patients.

Cross-species AI foundation model enables precision survival prediction in osteosarcoma

POSTER

179

Michelle White[1], Eddie Li[1]

1. Clyra Bio, Inc., San Francisco, CA, USA

Pediatric and adolescent osteosarcoma (OS) is exceedingly rare in humans (~500 US cases/year), constraining dataset sizes for prognostic modeling and clinical trial stratification. Conversely, spontaneous canine OS is roughly 20-fold more prevalent, offering an ecologically and therapeutically congruent model. To exploit this incidence asymmetry, we developed a cross-species deep learning framework that transforms canine data abundance into a clinical advantage: a genomic foundation-model survival predictor trained on canine bulk RNA-seq and engineered for direct zero-shot transfer to human patients.

Rather than quantifying gene-level expression against a species-specific reference annotation, our model embeds individual RNA-seq reads using Evo2, a sequence-to-function genomic foundation model pretrained across a broad range of species' DNA. Read-level Evo2 embeddings are aggregated using an attention mechanism that dynamically weights informative reads per sample. These aggregated sample representations are then passed to a survival-prediction head trained to predict time-to-event outcomes, such as disease-free interval and overall survival, directly from canine cases, then evaluated for transfer to human osteosarcoma cohorts.

Because the model operates on sequence embeddings rather than curated, ortholog-mapped gene sets, it captures shared underlying disease structure across species without requiring an a priori choice of which genes or pathways to examine, and without depending on cell-type deconvolution against a reference expression panel, a step that is itself species- and reference-dependent and can obscure patient-level heterogeneity. This yields a continuous, individualized survival estimate per patient, rather than a coarser prediction anchored to a discrete molecular subtype or immune cell-type composition.

This approach offers two direct benefits for human oncology: first, it allows canine osteosarcoma, where sample sizes are an order of magnitude larger, to substitute for scarce human training data; second, it produces a patient-level prognostic score usable for stratifying enrollment in human clinical trials, potentially improving trial power and success rates by enriching for patients most likely to benefit. We test the model on an independent dataset of human osteosarcoma patients, where its predictions are strongly predictive of developing metastasis within five years.

Superior accuracy and resolution for HLA typing from short-read sequencing with Mako

POSTER

180

Liza Huijse[1], Mary Wood[1], Solomon Adams[1], Julianne David[1], Solomon Endlich[1]

1. Benthic Genomics, Inc., Palo Alto, CA, USA

Human Leukocyte Antigen (HLA) genes are among the most highly polymorphic in the human genome and are critical determinants of immune function, playing important roles in precision health applications such as immunotherapy, transplantation, and disease risk assessment. Though next generation sequencing (NGS) data is widely available, accurate HLA typing remains challenging due to the structural complexity of the HLA locus. We present Mako, a pangenome graph-based genomics analysis system, which provides highly accurate HLA typing from NGS data and outperforms widely used methods.

Mako predicts alleles for 45 genes, exceeding T1K and DRAGEN HLA (41), HLA-HD (29), SpecHLA (8), and OptiType (3). Mako also calls alleles at up to 4-field resolution, exceeding the 3-field resolution of T1K and HLA-HD and 2-field resolution of OptiType. In addition to HLA typing, Mako generates variant calls, achieving performance on par with DeepVariant and DRAGEN against the NIST v4.2.1 benchmark, as well as phased diploid MHC assemblies, offering greater sensitivity for polymorphic regions.

We evaluated Mako's HLA calling on Illumina paired-end sequencing across 124 samples using pb-StarPhase calls on long-read sequencing as truth for 11 genes. At 2-field resolution, Mako achieved 98.5% concordance, outperforming HLA-HD (95.1%), T1K (93.3%), and DRAGEN HLA (88.3%); concordance remained higher on genes shared with SpecHLA (99.4% vs. 94.2%) and Optitype (99.9% vs. 97.2%). At 3-field resolution, Mako achieved 97.9% concordance, exceeding HLA-HD (93.5%), T1K (92.6%), and DRAGEN HLA (86.3%), and maintaining higher concordance than SpecHLA on shared genes (98.4% vs. 90.8%). We further benchmarked all methods using high-fidelity phased assemblies across a broader set of HLA genes, where Mako consistently achieved the highest accuracy, exceeding the other methods by up to 10.2% at 2-field resolution and 11.0% at 3-field resolution.

These findings demonstrate that pangenome graph-based assembly and annotation can overcome the limitations of reference-guided HLA typing approaches, enabling more comprehensive and accurate characterization of the HLA locus from existing NGS datasets. By enabling high-resolution HLA typing without specialized assays, Mako provides a scalable solution for precision health applications in immunology.

POSTER

181

Upwell target enrichment: a high-molecular-weight, methylation-preserving enrichment kit delivers full-resolution immunogenomics of the MHC and LRC

Solomon Endlich[1], Russell S. Julian[1], Vanessa B. Tran[1], Julianne K. David[1], Solomon M. Adams[1], Mary A. Wood[1], Janine Mok[1], Liza Huijse[1]

1. Benthic Genomics, Inc., Palo Alto, CA, USA

Background: The major histocompatibility complex (MHC) and leukocyte receptor complex (LRC) are central to immune function, making them critical to disease risk and therapeutic response. They are also among the most challenging regions of the human genome due to extreme allelic diversity, segmental duplication, copy-number variation, and extended linkage disequilibrium. Existing targeted enrichment approaches fragment DNA, disrupt long-range phase, or remove native DNA modifications, limiting comprehensive long-read immunogenomic analysis. We developed Upwell Target Enrichment, a Cas-based workflow that preserves long, unamplified, modification-bearing DNA for fully phased analysis of complex immune loci.

Methods: Upwell Target Enrichment protects target regions while degrading off-target DNA, preserving double-stranded, high-molecular-weight DNA with native modifications. The panel targets the complete LRC (all KIR and LILR genes), the MHC (HLA class I and II genes, MICA, MICB, TAP1, TAP2, the C4 region, and flanking sequence), and the CYP3A cluster. Performance was evaluated using whole-blood samples and reference cell lines (HG00514, HG00733) by short-read sequencing, digital PCR (dPCR), methylation-aware TAPS+ libraries, and fragment size analysis. To complement these, long-read Oxford Nanopore (ONT) sequencing datasets on a single blood donor were generated.

Results: Across four blood donors and two reference cell lines, Upwell achieved robust enrichment across all target loci (>15× sequencing coverage) while maintaining uniform coverage relative to whole-genome controls. Independent dPCR measurements typically demonstrated >30× enrichment. Enrichment was retained in methylation-aware (TAPS+) libraries, with strong signal across the classical HLA class II region (HLA-DR/DQ/DP). Similarly, ONT datasets show robust enrichment of target loci (>15x target enrichment over the background) while maintaining long-read lengths (majority of targeted coverage coming from reads >10kb), demonstrating its suitability for long-read sequencing.

Conclusions: A single Cas-based workflow delivers uniform enrichment of the MHC, LRC, and CYP3A while preserving unamplified, high-molecular-weight DNA and native DNA modifications. Upwell Target Enrichment provides the substrate for fully phased, methylation-aware long-read immunogenomic analysis.

From brain tissue to nanopore sequencing with minimal hands-on time

POSTER

182

Sebastian Rammensee[1], Simon Lesbirel[1], Kate Rouse[1], Simon Deacon[2,3], Simon Paine[3], Matthew Loose[2]

1. TTP Plc, TTP Campus, Melbourn, SG8 6HQ, UK
2. Deep Seq, Centre for Genetics and Genomics, The University of Nottingham, Queen's Medical Centre, Nottingham, NG7 2UH, UK
3. Nottingham University Hospitals NHS Trust, Queen's Medical Centre, Nottingham, NG7 2UH, UK

Background: Rapid sequencing has the potential to move genomic information closer to the point of clinical decision-making, but sample preparation remains a major barrier. Existing workflows typically require trained operators, multiple instruments and repeated manual liquid-handling steps, thereby limiting reproducibility, throughput and deployment outside specialist laboratories. This challenge is particularly acute for time-sensitive applications such as intraoperative molecular classification of brain tumours.

Methods: We are developing an automated sample-preparation workflow for Oxford Nanopore (ONT) sequencing using Puck, TTP's clinically proven, compact, low-cost "robot in a coffee pod" platform. Puck uses a disposable robotic pipette cartridge to perform programmable laboratory operations including reagent transfer, mixing, incubation, magnetic-bead processing and elution. The sequencing preparation workflow integrates DNA extraction from tissue with rapid nanopore library preparation. Feasibility was evaluated using pig brain tissue as a representative complex solid-tissue matrix. Extracted DNA and sequencing libraries were assessed for yield, purity, processing time, sequencing output and read-length distribution. To demonstrate relevance to a clinical workflow, sequencing data were analysed using ROBIN, a real-time nanopore analysis framework developed at the University of Nottingham for brain-tumour methylation classification and molecular profiling.

Results: The automated workflow successfully completed DNA extraction and sequencing-library preparation from pig brain samples in **under 2 hours**. DNA concentration ranged from **90-250 ng/ul with fragment lengths between 5-30 kb**, while sequencing generated a read N50 of **5-15 kb**. This performance is essentially equivalent to a trained operator, with hands-on time reduced from 1 hour for the manual workflow to 5 minutes for the automated workflow.

Conclusions: These feasibility data demonstrate that complex tissue-to-sequencer preparation can be transferred onto a compact, single-use automation platform. By reducing operator dependence and instrument burden, Puck could enable reproducible nanopore sequencing in decentralised laboratories and time-critical clinical settings. The integration of multiple workflow steps into a single instrument could also substantially simplify governance and operational oversight in clinical settings compared with the use of multiple separate pieces of equipment. Future work will optimise performance with human brain-tumour samples and assess end-to-end integration with the Nottingham ROBIN workflow.

Diagnostic yield of genome-based newborn screening in critically ill children: a retrospective analysis

POSTER

183

Erica Sanford Kobayashi[1,2,3], Laura Tobin[1], Edwin Juarez Rosales[1], Ofelia Vargas Shiraishi[2], Jason Knight[2], Jerica Lenberg[1], Rebecca Reimers[1], Kristen Wigby[1,3], Stephen F. Kingsmore[4], Matthew Bainbridge[1]

1. Rady Children's Institute for Genomic Medicine, San Diego, CA, USA
2. Children's Hospital Orange County, Orange, CA, USA
3. Rady Children's Hospital San Diego, San Diego, CA, USA
4. Cleveland Clinic, Cleveland, OH, USA

Over the past ten years, rapid diagnostic genomic sequencing (RDGS) has become increasingly utilized in critically ill children with illnesses of unclear etiology. However, mortality rates in some reported cohorts have exceeded 20%, prompting growing interest in identifying affected children before their first critical care admission. Population genome-based newborn screening (gNBS) is currently being investigated in several large prospective clinical studies as an adjunct to conventional newborn screening (NBS), with the goal of detecting childhood-onset genetic conditions before clinical symptoms develop and before associated morbidity and mortality occur.

To determine the proportion of ICU patients with genetic disorders that might have been detected through population gNBS at birth, we conducted a multicenter retrospective observational cohort study. The study included critically ill infants and children admitted to neonatal, pediatric, or cardiac intensive care units at two tertiary pediatric hospitals (CHOC and Rady Children's) who underwent RDGS as part of routine clinical care during their ICU stay. RDGS findings were translated into hypothetical gNBS results using 778 genes included across three established gNBS platforms.

Among 1,044 infants and children who underwent RDGS during an ICU admission, 29% received reportable likely pathogenic or pathogenic findings. Overall, 11% of all patients tested by RDGS would have been expected to screen positive through gNBS, accounting for 39% of those with a diagnostic genetic result. Furthermore, 87% of patients with likely pathogenic or pathogenic variants that would have been detectable by gNBS experienced a documented change in clinical management following receipt of their RDGS diagnosis.

These results indicate that population gNBS could have a substantial effect on children who later require admission to neonatal, pediatric, or cardiovascular ICUs. In cases where ICU admission or RDGS testing was delayed, gNBS may enable earlier diagnosis and potentially reduce the risk of progression to critical illness. Collectively, these findings provide additional support for population gNBS and highlight its potential to decrease ICU admissions and improve clinical outcomes.

POSTER

184

truCOLLECT whole blood collection and truCOVER library prep solution: from decentralized sample collection to pharmacogenetic profiling

George Endress[1], Heidi Giese[1], Ulrich Thomann[1], Elinette Albino Rodriguez[3], Vanessa Process[1], Paola Giusti Rodriguez[2], Kelly Dabrowski[1], Jose Gil[1], Kasra Tabatabei[1], Eugenio Daviso[1]

1. Covaris LLC, a PerkinElmer Company, Woburn, MA, USA
2. University of Florida, Department of Psychiatry, Gainesville, FL, USA
3. University of Puerto Rico, Medical Campus, San Juan, Puerto Rico, USA

The collection of high-quality genomic DNA remains a major barrier in pediatric, and especially in neurodevelopmental research, particularly among children with autism spectrum disorder (ASD) and other neurogenetic conditions, where venous blood draws may be distressing or logistically challenging. Additionally, access to phlebotomy services can be limited in rural and remote settings, creating disparities in genomic research participation.

We present results from a pilot study evaluating truCOLLECT fingerstick-based blood collection kits as an alternative to traditional venous blood draws. 50 µl of fingerstick blood were collected, immobilized and stabilized, allowing storage and transport at ambient temperature without cold-chain or biohazard logistics. Collection settings were in rural sites in Puerto Rico. DNA was extracted up to 6 months after collection, with no evidence of degradation. Libraries were constructed (truCOVER WGS PCR-free Library Prep) and sequenced (NovaSeq X). Data analysis (PharmCAT and StellarPGx) was focused on pharmacogenetic markers and population genetics (brentp/somalier).

The truCOLLECT Solution showed significant practical advantages over phlebotomy- or saliva-based sampling. Phenotypic concordance of all analyzed pharmacogenetic loci between truCOLLECT and phlebotomy samples was 100%. Success rates of truCOLLECT Solution-based sample collection/storage and shipment, DNA extraction and NGS data analysis were 98%, 100% and 100%, respectively.

From consent to continuity: a participant-centered framework for translational precision health

POSTER

185

Kaya Bilguvar[1,2,3,5,6,7,8,9,10,11,12], Aybike S. Bulut[1], Merve Nur Koroglu Karakaya[2], Ozkan Ozdemir[1,3,4,5], Murat Gunel[6,7,8,9,10,11]

1. Acıbadem Mehmet Ali Aydınlar University, Health Sciences Institute, Department of Translational Medicine, Istanbul, Turkiye
2. Acıbadem Mehmet Ali Aydınlar University, Health Sciences Institute, Department of Biostatistics and Bioinformatics, Istanbul, Turkiye
3. Acıbadem Mehmet Ali Aydınlar University, Health Sciences Institute, Department of Genome Studies, Istanbul, Turkiye
4. Acıbadem Mehmet Ali Aydınlar University, School of Medicine, Department of Medical Biology, Istanbul, Turkiye
5. Acıbadem Mehmet Ali Aydınlar University, ACURARE - Rare Diseases and Orphan Drugs Application and Research Center, Istanbul, Turkiye
6. Yale School of Medicine, Department of Neurosurgery, New Haven, CT, USA
7. Yale School of Medicine, Department of Neuroscience, New Haven, CT, USA
8. Yale School of Medicine, Yale Program on Neurogenetics, New Haven, CT, USA
9. Yale Center for Genome Analysis, Yale School of Medicine, New Haven, CT, USA
10. Yale School of Medicine, Department of Genetics, New Haven, CT, USA
11. Yale School of Medicine, Yale Program in Brain Tumor Research, New Haven, CT, USA
12. Acıbadem Mehmet Ali Aydınlar University, School of Medicine, Department of Medical Genetics, Istanbul, Turkiye

Precision health has fundamentally transformed how research participants contribute to biomedical discovery. A single participant's contribution may generate multi-omic datasets, patient-derived cell lines, induced pluripotent stem cells (iPSCs), organoids, artificial intelligence (AI)-assisted analyses, and future translational applications extending decades beyond the original study. However, governance remains largely organized around individual research activities including informed consent, biospecimen collection, ethics review and data sharing rather than the evolving relationship between participants and translational research. Drawing on complementary experience from the establishment of a participant-centered precision health infrastructure at Acıbadem Mehmet Ali Aydınlar University (approximately 307 participants recruited over five years across multiple rare disease cohorts) together with governance insights from a long-standing neurogenetics research program comprising more than 4,200 whole-exome sequencing datasets collected over two decades across Türkiye and the United States, we examined how governance requirements evolve throughout translational research. We observed that participant-derived scientific resources consistently outlive the studies for which they were originally collected. Biospecimens initially obtained for diagnostic or disease-specific research frequently generate new multi-omic datasets, patient-derived cellular models, iPSCs, organoids, and future translational applications that could not reasonably have been anticipated at enrollment. As technologies advance, these resources continue to acquire scientific value through reanalysis, international collaboration, controlled-access repositories, and AI-assisted analyses. These observations suggest that governance challenges commonly treated as independent issues including consent, biospecimen stewardship, secondary use, data sharing, participant recontact and return of findings are better understood as interconnected consequences of a single longitudinal participant–research relationship. We present a participant-centered continuity framework that shifts governance from isolated regulatory events toward continuous stewardship of participant-derived scientific resources, providing a conceptual foundation for next-generation translational precision health.

POSTER

186

Cryopreserved tumor biospecimens improve whole-genome sequencing performance and precision oncology variant detection compared with formalin-fixed tissue

Diana Nortey-Annan[1], Ken Dixon[2], Jeffrey Okojie[1,3], Jared Barrott[1,2,3,4]

1. Brigham Young University, Cell Biology and Physiology, Provo, UT, USA
2. SpeciCare, Gainesville, GA, USA
3. Brigham Young University, Simmons Center for Cancer Research, Provo, UT, USA
4. Brigham Young University, School of Medicine, Provo, UT, USA

Pre-analytical biospecimen preservation is a critical but often overlooked determinant of genomic data quality in precision oncology. Although formalin-fixed paraffin-embedded (FFPE) tissue remains the standard clinical specimen, formalin fixation introduces DNA fragmentation and sequencing artifacts that may compromise molecular analyses. We evaluated whether cryopreserved (CP) tissue improves DNA integrity and downstream whole-genome sequencing (WGS) performance using matched pan-cancer tumor specimens.

Matched CP and FFPE tumor samples representing diverse cancer histologies were analyzed across two complementary studies. DNA yield, purity, fragment integrity, and high-molecular-weight DNA content were quantified in 38 matched tumor pairs. WGS was subsequently performed on an independent cohort of 50 matched tumor pairs to compare sequencing quality metrics, variant concordance, tumor mutational burden, structural variant detection, and clinically relevant cancer gene identification.

Cryopreserved tissue consistently produced higher-quality genomic DNA than FFPE tissue, yielding over four-fold greater DNA recovery, substantially higher DNA integrity scores, and approximately nine-fold more DNA fragments exceeding 40 kb in length. These improvements translated directly into superior WGS performance, including significantly greater read depth, larger insert sizes, reduced discordant read rates, and improved sequencing quality metrics. Across matched tumor pairs, variant concordance between preservation methods was only 43.5% at conventional variant allele frequency thresholds, while FFPE samples demonstrated inflated tumor mutational burden and increased multinucleotide variants and indels consistent with fixation-associated artifacts. Cryopreserved tissue also identified more structural variants and clinically relevant cancer gene alterations, including additional oncogenic driver mutations that were absent from matched FFPE analyses.

These complementary studies demonstrate that biospecimen preservation directly influences genomic accuracy from DNA extraction through clinical variant interpretation. While FFPE remains indispensable for histopathology, cryopreserved tissue preserves native DNA integrity, improves whole-genome sequencing performance, and enhances detection of clinically actionable genomic alterations. Integrating cryopreservation into precision oncology workflows could substantially improve the reliability of comprehensive genomic profiling and emerging long-read sequencing applications.

POSTER

187

Long-read single-cell transcriptomics nominates predicted endogenous dsRNA-forming isoforms and maps state-resolved burden in colorectal cancer

Alex Fang[1], Yunyun Zhou[2]

1. Independent Researcher, Irvine, CA, USA
2. Fox Chase Cancer Center, Philadelphia, PA, USA

Long-read single-cell RNA sequencing resolves full-length transcript structures that gene-level analyses cannot preserve. We tested whether dsRNA-relevant transcript features—sequence complementarity and inverted Alu architecture—could nominate endogenous double-stranded RNA (dsRNA)-forming potential in colorectal cancer (CRC) and quantify its distribution across malignant epithelial states.

We analyzed 3,262 epithelial cells from 24 matched tumor/adjacent-normal samples across 12 CRC patients and curated 26,091 FSM/ISM isoforms. Isoforms were scored using intra-isoform complementarity (S1) and inverted-Alu architecture (S2), with reliability required for eligibility (S3 ≥ 0.10). The active score, 0.863539 × S1 + 0.136461 × S2, nominated a rank-defined set of 76 isoforms from 70 loci. Local ±10% weight perturbations retained 75–76/76 candidates; S1-only and S2-only rankings retained 61/76 and 38/76, respectively, indicating local stability but dependence on both structural components. Against 901 J2-RIP-seq-positive and 901 matched negative loci, out-of-fold AUROC was 0.731 and average precision was 0.741; adding S2 improved AUROC by 0.0239 over S1 alone. Among candidate loci, 51/70 were represented in the J2 locus table; within the source-specific 62-gene set, 47/62 showed positive enrichment and 31/62 reached log2 enrichment ≥1.

Predicted candidate burden was higher in enterocyte-like iCMS3-like malignant cells (CEACAM5+/PHGDH+; median 23.5) than stem-like iCMS2-like cells (ASCL2+/AXIN2+; median 11), with concordant direction in 12/12 patients (one-sided binomial p = 0.000244). Malignant epithelial cells also exceeded adjacent-normal epithelium (median 13 vs 9; 11/12 patients; p = 0.00317). Cross-fitted patient-aware validation retained positive normalized and raw burden differences in 11/12 patients, but the prespecified detectability-matched null did not establish candidate-specific enrichment (normalized empirical p = 0.154; raw p = 0.357).

CRC biochemical data provided locus-level support. MDA5 RNase-A-protected RNA was enriched at candidate loci in HT29 and HCT116 cells (empirical p < 0.0001 in each), with strongest localization to 3′-UTR and Alu-arm regions. Long-read CRC data reproduced exact intron chains for 46/76 candidates but showed no enrichment over matched controls (p = 0.267), and a preregistered multi-cohort state sweep did not independently reproduce candidate-specific differentiation/iCMS enrichment.

The framework therefore supports stable isoform nomination and a patient-consistent within-cohort state association while distinguishing structural prediction from independent replication, physical dsRNA formation, sensor activation, and clinical utility.

Index

	Last Name	First Name	Organization / Company	Page
A	Adams	April	Baylor College of Medicine	40
	Amini	Hajar	University of California, Davis	63
B	Baya	Nikolas	Mass General Brigham	46
	Bilgüvar	Kaya	Acibadem Mehmet Ali Aydınlar University	102
	Blankson	Harriet	Morehouse School of Medicine	54
	Blucher	Corrine	Rady Children's Institute for Genomic Medicine	25
	Bodkin	Nicholas	Northwestern University	34
	Bowman	Derek	Michigan State University	83
	Brashear	Awtum	Shriners Children's	95
	Broad	Mia	Northwestern University	74
	Brokamp	Elly	Boston Children's Hospital	82
	Bush	William	Case Western Reserve University School of Medicine	93
C	Caggiano	Christa	Icahn School of Medicine at Mount Sinai	42
	Calame	Daniel	Baylor College of Medicine	89
	Carter	Javan	RTI international	50
	Chen	Han	University of California, San Diego	51
	Chen	Chi-Yuan	Chang Gung University of Science and Technology	55
	Cheng	Yong-Han Hank	University of Washington	61
	Cohen	Tom	Panome Bio	94
	Cooperstein	Isabelle	University of Utah	71
D	DiTroia	Stephanie	Broad Institute of MIT and Harvard	45
	Dutka	Tara	National Institutes of Health	26
E	Endlich	Solomon	Benthic Genomics	98
	Endress	George	Covaris, LLC.	101
	Evanich	Daniel	New England Biolabs	23
	Ewool	Kenneth	Brigham Young University	48
F	Fang	Alex	Independent	104
G	Gallagher	Michael	Hitachi	33
	Gibson	Sophia	University of Washington	60
	Goddard	Page	Rady Children's Institute for Genomic Medicine	31
	Gray	Kathryn	University of Washington	80
	Gupta	Pankhuri	University of Washington	87
	Gyamfi	Angela	St Jude Children's Research Hospital	37

Index

	Last Name	First Name	Organization / Company	Page
H	Holuka	Cyrielle	St Jude Children's Research Hospital	57
	Hooker	Qualia	Van Andel Research Institute	56
	Hujise	Liza	Benthic Genomics	97
J	Jones	Ewin	Cell Biology and Physiology Brigham Young University	44
K	Kennedy	Shariya	Omega Bio-tek	59
	Kokaras	Alexis	New England Biolabs	85
	Kosicki	Michael	Arc Institute / LBNL	66
	Kuufire	Emmanuel	Tuskegee University	62
L	Larkin	Katie	Broad Clinical Labs	28
	Li	Fuchen	Johns Hopkins University	79
	Li	Mingyuan	Johns Hopkins University	29
	Li	Qiuhui	Johns Hopkins University	75
	Li	Yaqiao	Gladstone Institutes	30
	Linder	Jodell	Vanderbilt University Medical Center	24
M	Maguire	Anne	Auburn University	92
	Mahinparvar	Nazanin	Stanford University	68
	Major-Mincer	Joshua	University of Utah	43
	McVicker	Graham	Salk Institute for Biological Studies	58
	Metcalf	Ginger	Baylor College of Medicine	52
	Mojica	Wilfrido	Virchow Medical, Inc	77
N	Nguyen	Quang Huy	Auburn University	70
	Nortey-Annan	Diana	Brigham Young University- Provo	103
O	Olszewski	Athena	St. Jude Children's Research Hospital	91
P	Pandey	Devansh	The University of Texas at Austin	32
	Pardo	Katherine	Undiagnosed Diseases Program, National Institutes of Health	49
R	Rammensee	Sebastian	TTP	99
	Reimers	Rebecca	Rady Children's Institute for Genomic Medicine	90

Index

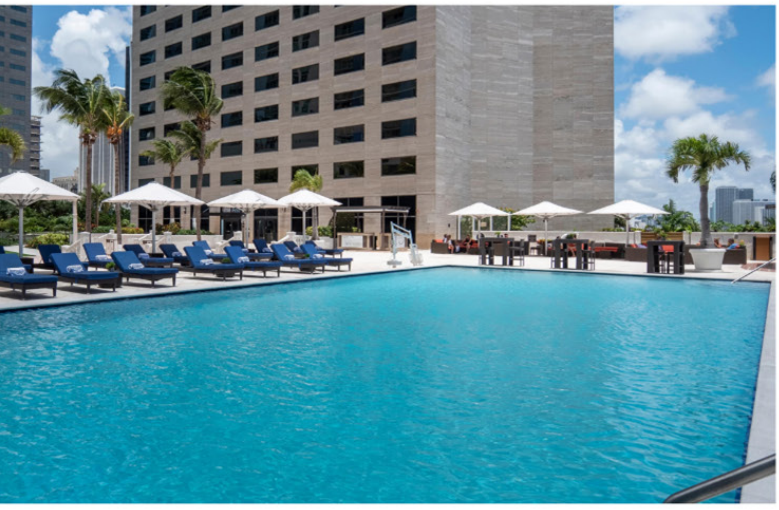
	Last Name	First Name	Organization / Company	Page
S	Salz	Lisa	Baylor Genetics	39
	Sanford Kobayashi	Erica	Rady Children's Institute for Genomic Medicine	100
	Sato	Josei	UCSF	47
	Setia-Verma	Shefali	University of Pennsylvania	20
	Shah	Saanchi	Florida Institute for Pediatric Rare Diseases, Florida State University College of Medicine	86
	Slaven	Neil	Lawrence Berkeley National Lab	81
	Steveson	Isaac	Barrott Lab	41
	Strickland	Gwendolyn	Boston Children's Hospital	53
	Sweeney	Michael	University of Michigan	76
T	Tawa	Gregory	NIH/NCATS/TDB	22
	Thompson	Whitney	Mayo Clinic Children's	88
	Todeschini	Andrew	Brigham Young University	65
	Tran	Stephen	UCSD	73
	Tran	Thien	UCSF	38
	Tsai	Dianrong	BYU	67
	Tuncel	Tunç	Health Institutes of Türkiye, Türkiye Biotechnology Institute	35
	Tunstall	Tate	MyOme	36
V	Vaz	Eduarda	Johns Hopkins University	72
	Verma	Rashi	Morehouse School of Medicine	27
W	Wang	Kanix	University of Cincinnati	84
	Warnick	Caden	Brigham Young University, Cell Biology and Physiology	21
	White	Michelle	Clyrabio	96
X	Xie	Yilin	Stanford University	64
Y	Yalcouye	Abdoulaye	Johns Hopkins University	69
Z	Zhao	Chunyu	CZ Biohub SF	78

AGBT 2027™
AGRICULTURE

SAVE THE DATE

April 4-7, 2027

InterContinental Miami



Mark your calendar and learn more at www.agbt.org

MilliporeSigma

Together for **Life.** With **Science.**

From the first spark of discovery to full-scale manufacturing, **MilliporeSigma** stands beside you. As a global leader in Life Science, we offer over 300,000 products and comprehensive solutions: from everyday lab essentials to Gold Standard water purification systems, from GMP-grade manufactured raw materials to customized bioprocessing systems, from large-scale ultrafiltration devices to contract testing and CDMO services.

Everything you need. One trusted partner.

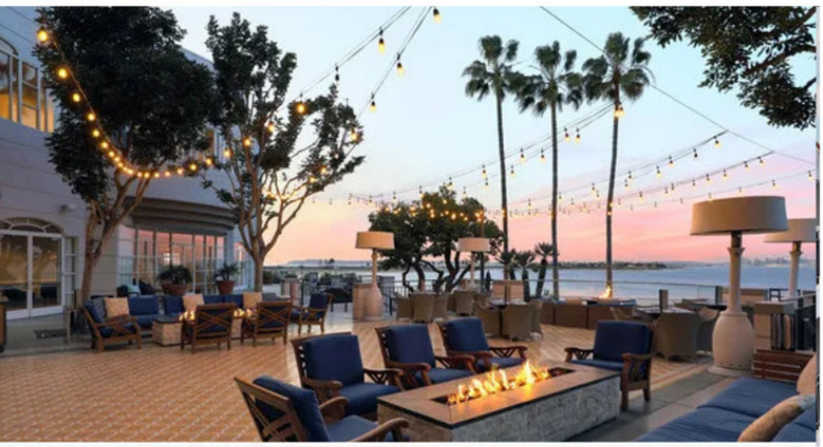
www.milliporesigma.com

AGBT2027™
PRECISION HEALTH

SAVE THE DATE

SEPTEMBER 13-15, 2027

Loews Coronado Bay Resort
San Diego, CA



Mark your calendar
and learn more at
WWW.AGBT.ORG