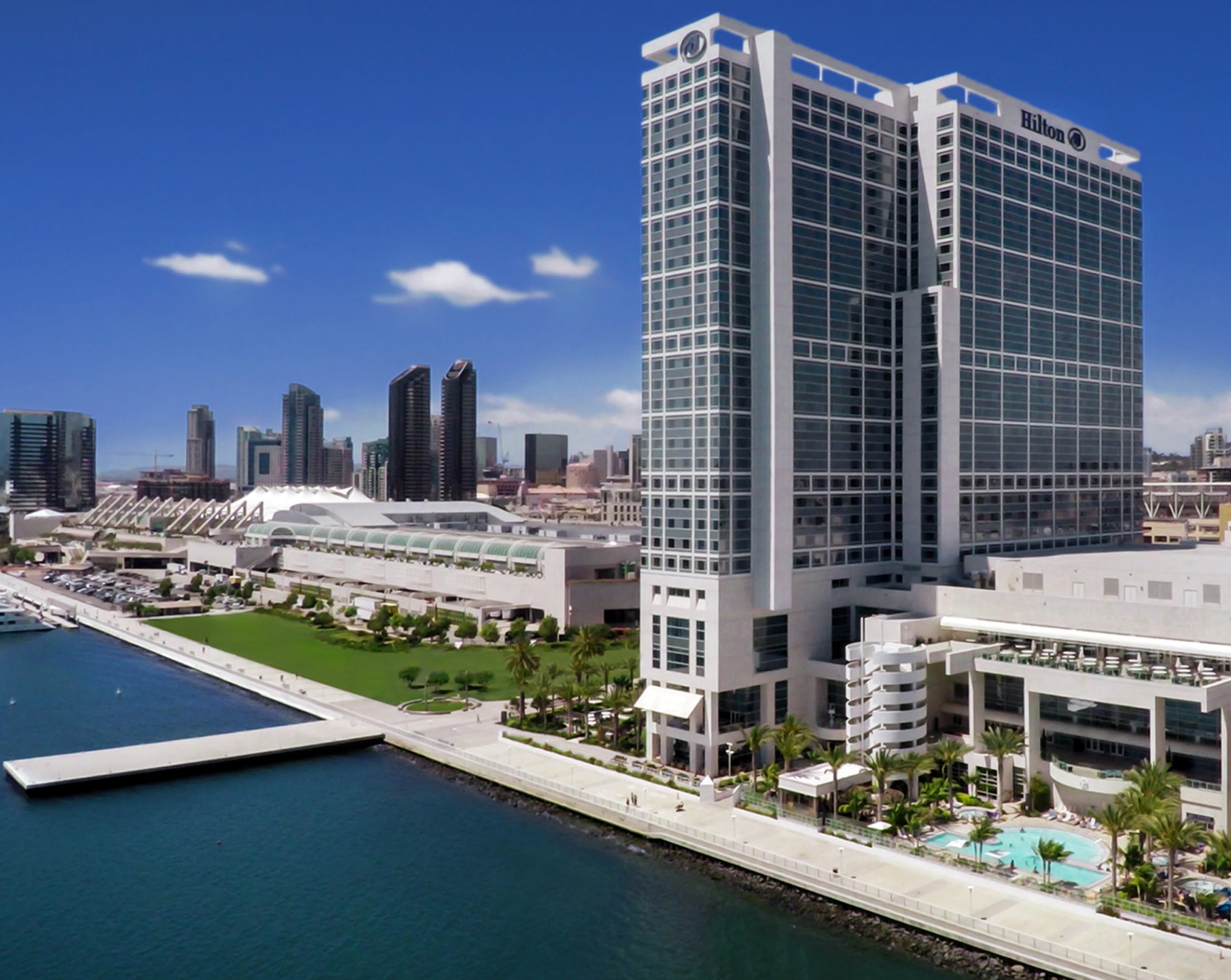
AGBT2023™

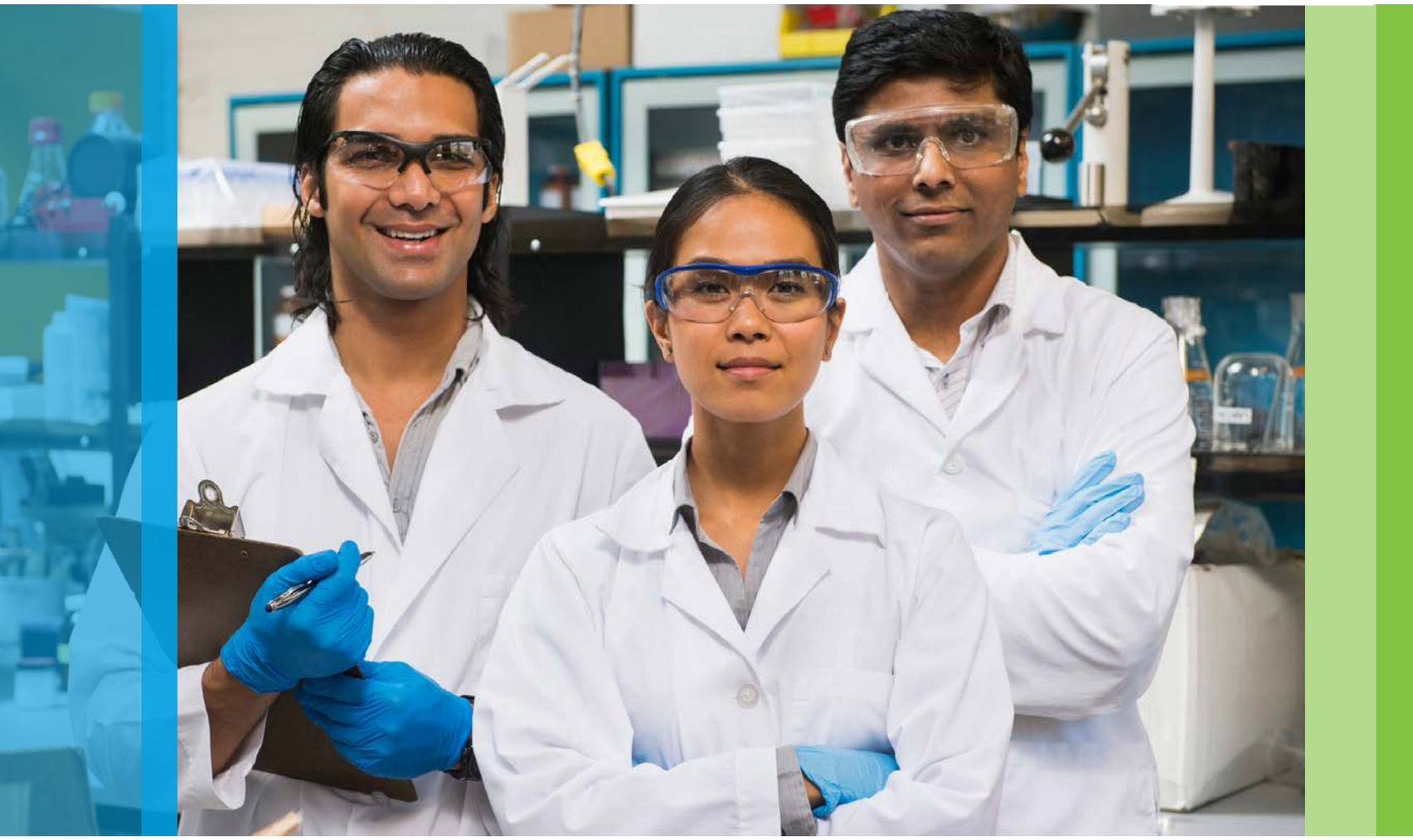
PRECISION HEALTH

September 7-9 • San Diego, CA



Agilent

Trusted Answers



Your Partner for Delivering Complete Genomics Workflow Solutions

Agilent is a global leader in the life sciences, diagnostics, and applied chemical markets, delivering insights and innovations to advance the quality of life. We offer a full range of high-quality workflow solutions within precision health, translational research, companion diagnostics and more. Find everything you need to support or expand your laboratory, including sample quality control (QC), next-generation sequencing (NGS) library prep and target enrichment, microarrays, CRISPR, PCR/qPCR, bioreagents, and data analysis solutions including interpretation, and reporting.

Learn more here

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

This information is subject to change without notice.

Visit Agilent @AGBT PH

Published in the USA, June 2023
PR7001-1272

AGBT 2023™
PRECISION HEALTH

It is our honor to welcome you to the 2023 Advances in Genome Biology and Technology (AGBT) Precision Health meeting.

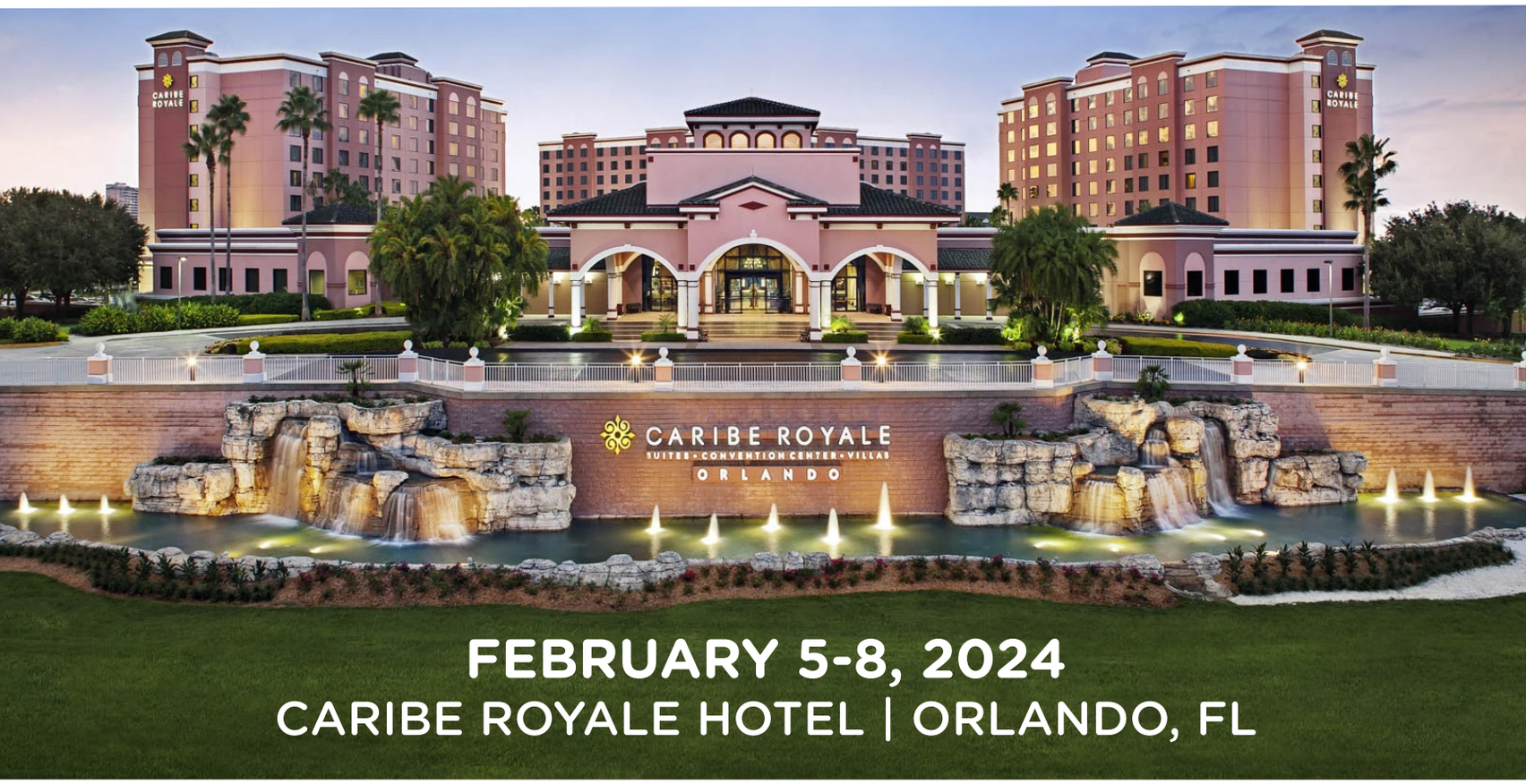
Now in its 8th year, AGBT is recognized as a premier gathering for the genomic research community. Since its inception, this meeting has highlighted the most provocative discoveries and major initiatives in genomic medicine from leaders of national healthcare systems, genomics institutes, international biobank initiatives, transformative diagnostic approaches, and hospital-based implementation and reimbursement.

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

AGBT is a not-for-profit organization providing three prestigious conferences and networking events. Our next meetings are The General Meeting (February 5-8, 2024) and AGBT Agriculture (April 15-17, 2024). Learn more at agbt.org.

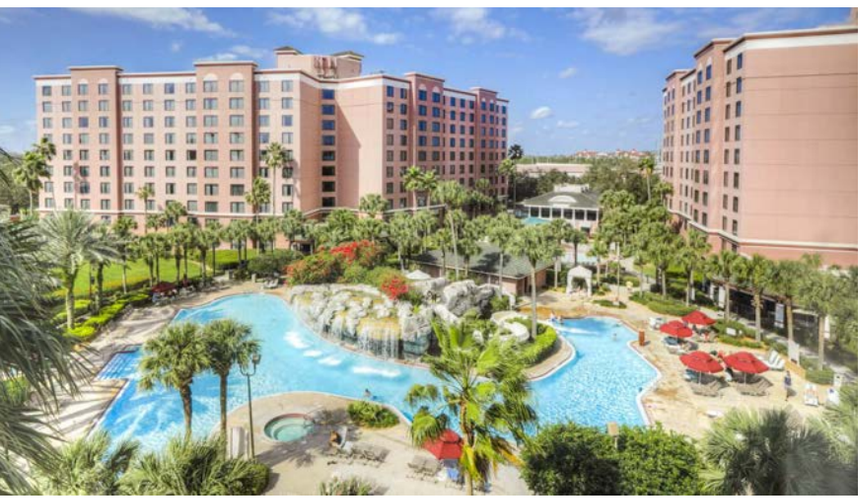
AGBT2024™

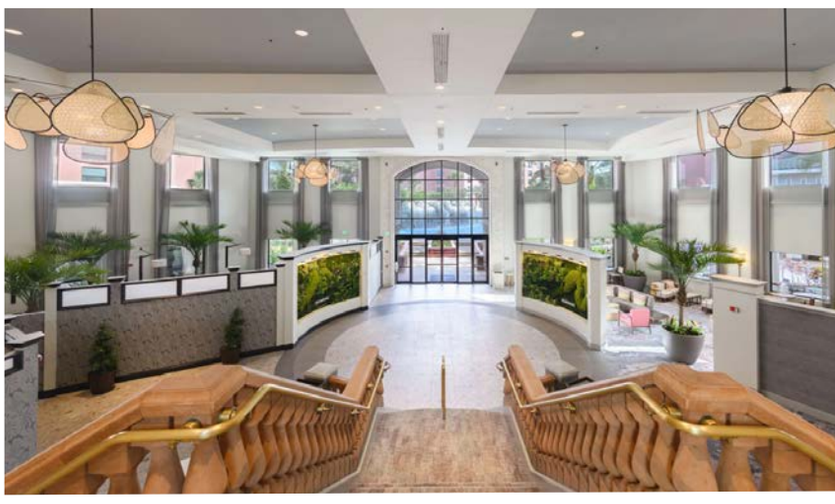
GENERAL MEETING



FEBRUARY 5-8, 2024

CARIBE ROYALE HOTEL | ORLANDO, FL





ABSTRACTS OPEN SEPT 6TH

SEE OUR SPEAKERS AND LEARN MORE BY SCANNING THE QR CODE, OR VISIT WWW.AGBT.ORG

Organizing Committee

We are incredibly grateful to our meeting Co-Chairs Eric Green, Wendy Chung, Mike Talkowski, and the entire Scientific Organizing Committee for the countless hours they invested in making AGBT Precision Health 2023 a content-rich experience.

Penelope Bonnen
Baylor College of Medicine

Wendy Chung
Columbia University Medical Center

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

Eric Green
National Human Genome Research Institute

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

Professor Dame Sue Hill
NHS, England

Stephen Kingsmore
Rady Children's Institute for Genomic Medicine

Howard Mcleod
Intermountain Healthcare

Len Pennacchio
Lawrence Berkeley National Laboratory, DOE
Joint Genome Institute

Kathryn Phillips
University of California, San Francisco

Heidi Rehm
The Broad Institute of MIT and Harvard

Mike Talkowski
The Broad Institute of MIT and Harvard

AGBT 2023™
PRECISION HEALTH

We gratefully acknowledge the generous support provided by the following companies

DIAMOND

Agilent

EMERALD

CODEXIS®

SAPPHIRE

cambridge consultants
Part of Capgemini Invent

RUBY 1

FABRIC GENOMICS™

RUBY 2

CELLECTA

EXHIBITORS & CONTRIBUTING SPONSORS

ARGONAUT MANUFACTURING SERVICES
bionano
CELEMICS
L7 INFORMATICS
Intelliseq
omega BIO-TEK
innovations in nucleic acid isolation
PacBio
paradigm4
SINGULAR GENOMICS
STANDARD BIOTOOLS
TWIST BIOSCIENCE
Vazyme

General Information

Registration / Hospitality Desk

Thursday, September 7	11:00 a.m. – 7:30 p.m.
Friday, September 8	7:30 a.m. – 7:30 p.m.
Saturday, September 9	7:30 a.m. – 5:00 p.m.

Name Badges

Please always wear your name badge & lanyard. Security will refuse entry to anyone not associated with the AGBT meeting.

Hotel Information

Rising above the San Diego Bay and steps from the Gaslamp Quarter and Petco Park, Hilton San Diego Bayfront is the signature SoCal resort. Enjoy a tailored, urban coastal experience with bright and unparalleled views in every guest room, relaxing amenities, and locally inspired dining.

Throughout your trip, you may be reached as follows:

Hilton San Diego Bayfront
1 Park Blvd
San Diego, CA 92101

Telephone: 619-564-3333

Social Events

Welcome Reception – **Thursday, September 7, from 7:30 p.m. – 10:00 p.m.** in the Promenade Plaza

AGBT Dinner – **Friday, September 8, from 6:45 p.m. – 9:00 p.m.** in the Promenade Plaza

Conference Technology

Download our official AGBT conference mobile app for easy access to the agenda, abstracts, and more. Check your email for the link to download the app or visit our registration desk for assistance.

Start Posting!

#AGBTPH

cambridge consultants
Part of Capgemini Invent

Making precision medicine work

COME TO OUR TALK

To capture too much, or too little, that is the question: making genomic data work in a clinical world

SPEAKER
Symon Cotton
Saturday September 9th
12.20-12.35

Meet our team
at BOOTH 201

Find out more
about Life Science
at Cambridge Consultants

Agenda

Thursday, September 7, 2023

11:00 a.m. – 7:30 p.m.	**Hospitality Desk** Indigo West Foyer
	Pre-Conference Workshops
1:00 p.m. – 2:30 p.m.	**Navigating Clinical Genomic Databases** **Heidi Rehm, Broad Institute of MIT and Harvard, Kaitlin Samocha, Center for Genomic Medicine, Mass General Hospital (Workshop Co-Chairs)**
2:45 p.m. – 4:15 p.m.	**Leveraging the *All of Us* Researcher Workbench to Advance Precision Health** Participants will learn about the extensive data available on the All of Us Researcher Workbench and through a guided tour, complete an example genomic analysis. **Alexander Bick, Vanderbilt**
4:15 p.m. – 5:00 p.m.	**Welcome Drink - Exhibit Hall** Indigo Ballroom ABEF
5:00 p.m. – 5:10 p.m.	**Welcome**
Session I:	**Landscape of Genomic Medicine** **(Chairs - Eric Green, National Human Genome Research Institute and Michael Talkowski, Massachusetts General Hospital and Broad Institute Chairs)** Indigo Ballroom CDGH
5:10 p.m. – 5:40 p.m.	**Elaine Mardis, Co-Executive Director of the Institute for Genomic Medicine at Nationwide Children's Hospital** "Precision diagnostics for pediatric cancer nationwide"
5:40 p.m. – 6:10 p.m.	**Todd Golub, Director and Founding Member, Broad Institute of MIT and Harvard** "Perspectives on cancer precision medicine"
6:10 p.m. – 6:40 p.m.	**Allyson Berent, Chief Science Officer, FAST** "The parents' journey through drug development: A story in Angelman Syndrome; making the impossible possible"

6:40 p.m. — 7:00 p.m.	**Andrew Stergachis (Abstract Selected), University of Washington School of Medicine** “Comprehensive, integrated long-read genomic, epigenomic, and transcriptomic characterization for solving the molecular basis of Mendelian conditions”
7:00 p.m. — 10:00 p.m.	**Welcome Reception** Promenade Plaza

Friday, September 8, 2023 (Morning)

7:30 a.m. — 7:30 p.m.	**Hospitality Desk** Indigo West Foyer
7:30 a.m. — 9:00 a.m.	**Breakfast** Indigo Terrace
Session II:	**Genomic Medicine I: Developmental and Clinical Genomics (Chairs - Wendy Chung, Columbia University and Stephen Kingsmore, Rady Children’s Hospital)** Indigo Ballroom CDGH
9:00 a.m. — 9:30 a.m.	**Yufeng Shen, Columbia University** "Genetic architecture of developmental disorders"
9:30 a.m. — 10:00 a.m.	**Lyn Chitty, University College London** "The impact of genomics on the prenatal diagnosis service in the English National Health Service"
10:00 a.m. — 10:30 a.m.	**Alexandre Reymond, UNIL, Universite de Lausanne & Past President, ESHG** "Blurring the boundaries between rare and common diseases"
10:30 a.m. — 10:50 a.m.	**Meredith Wright (Abstract Selected), Rady Children’s Institute for Genomic Medicine** “Performance of BeginNGS, an artificial intelligence-enabled genome sequencing system, for universal newborn screening, diagnosis, and precision medicine for 410 severe childhood genetic diseases in over 460,000 individuals”

10:50 a.m. — 11:15 a.m.	**Coffee Break/Exhibit Hall** Indigo Ballroom ABEF
Session III:	**The Great Debate-Genetics, Environment and Health (Chairs - Heidi Rehm, Massachusetts General Hospital and Broad Institute, Geoff Ginsburg, Chief Medical and Scientific Officer, *All of Us* Research Program)**
11:15 a.m. — 12:15 p.m.	**Panelists:** **Gary Miller, Environmental Health Sciences, Columbia University** **Philip Awadalla, Ontario Institute for Cancer Research (OICR)** **Alison Motsinger-Reif, National Institute of Environmental Health Sciences**

Friday, September 8, 2023 (Afternoon)

12:15 p.m. — 2:00 p.m.	**AGBT Lunch** Indigo Terrace
12:30 p.m. — 2:00 p.m.	**Rady Children's Institute for Genomic Medicine, Lunch and Learn** "Empowering precision care and improving outcomes in the NICU and PICU by ultra-rapid genome sequencing" Room 204 A&B
2:15 p.m. — 2:50 p.m	**Sponsor Workshops (Howard Mcleod, Intermountain Healthcare, Chair)** Indigo Ballroom CDGH
2:15 p.m. — 2:35 p.m.	**Diamond Sponsor – Agilent Technologies** **Bellal Moghis, MS, MBA, Director of Product Marketing & Strategy, Agilent's Diagnostics and Genomics Group** "Solving challenges in multiomic liquid biopsy testing using the new Agilent Avida dual-target enrichment solution"

2:35 p.m. — 2:50 p.m.	**Emerald Sponsor – Codexis** **Mathew Miller, PhD, Associate Director, Life Science Technology & Applications, Codexis** "An Engineered DNA ligase for highly efficient sample litigation in NGS library preparation"
Session IV:	**Towards a Complete Human Genome** **(Chairs - Eric Green, National Human Genome Research Institute and Michael Talkowski, Massachusetts General Hospital and Broad Institute Chairs)** Indigo Ballroom CDGH
2:50 p.m. — 3:20 p.m.	**Karen Miga, University of California, Santa Cruz** "Expanding studies of rare disease using scalable long-read sequencing"
3:20 p.m. — 3:50 p.m.	**Michael Schatz, Johns Hopkins University** "Analysis of pancreatic cancer risk variants using long-read sequencing"
3:50 p.m. — 4:20 p.m.	**Danny Miller, University of Washington** "Streamlining clinical genetic testing with long-read sequencing"
4:20 p.m. — 4:40 p.m.	**Kiran Garimella (Abstract Selected), Broad Institute of MIT and Harvard** “New insights into genomic variation from long-read sequencing of >1000 African-American participants in the All of Us Research Program”
4:40 p.m. — 6:45 p.m.	**Poster Session & Exhibit Hall, Wine and Cheese Reception** Indigo Ballroom ABEF
6:45 p.m. — 9:00 p.m.	**AGBT Dinner** Promenade Plaza

Saturday, September 9, 2023 (Morning)

7:30 a.m. — 5:00 p.m.	**Hospitality Desk** Indigo West Foyer

7:30 a.m. — 9:00 a.m.	**Breakfast** Indigo Terrace
Session V:	**Session V: Genomic Mechanisms to Therapeutics (Chairs - Len Pennacchio, Lawrence Berkeley National Laboratory and DOE Joint Genome Institute, and Penelope Bonnen, Baylor University)** Indigo Ballroom CDGH
9:00 a.m. — 9:30 a.m.	**Melina Claussnitzer, Massachusetts General Hospital and Broad Institute of MIT and Harvard** "Translating metabolic genetic risk to cellular function"
9:30 a.m. — 10:00 a.m.	**Lea Starita, University of Washington and the Co-director of Brotman Baty Advanced Technology Lab** “Understanding the functional effects of genetic variation at scale”
10:00 a.m. — 10:30 a.m.	**Joe Gleeson, University of California, San Diego** "Leveraging patient cohorts with pediatric brain disease for discovery and therapy"
10:30 a.m. — 10:50 a.m.	**Altuna Akalin (Abtract Selected), Max Delbrueck Center** "Cardiovascular disease biomarkers derived from curculating cell-free DNA methylation"
10:50 a.m. — 11:20 a.m.	**Coffee Break & Exhibit Hall** Indigo Ballroom ABEF
Session Vl:	**Fireside Chat – a patient’s perspective with Wendy Chung, Columbia University** Indigo Ballroom CDGH
11:20 a.m. — 12:20 p.m.	**Luke Rosen, KIF1A.ORG, Founder** "Susannah's Story"
12:20 p.m. — 12:35 p.m.	**Sapphire Sponsor – Cambridge Consultants** **Symon Cotton, PhD, SVP of Lifescience, Cambridge Consultants** "To capture too much, or too little, that is the question: making genomic data work in a clinical world"

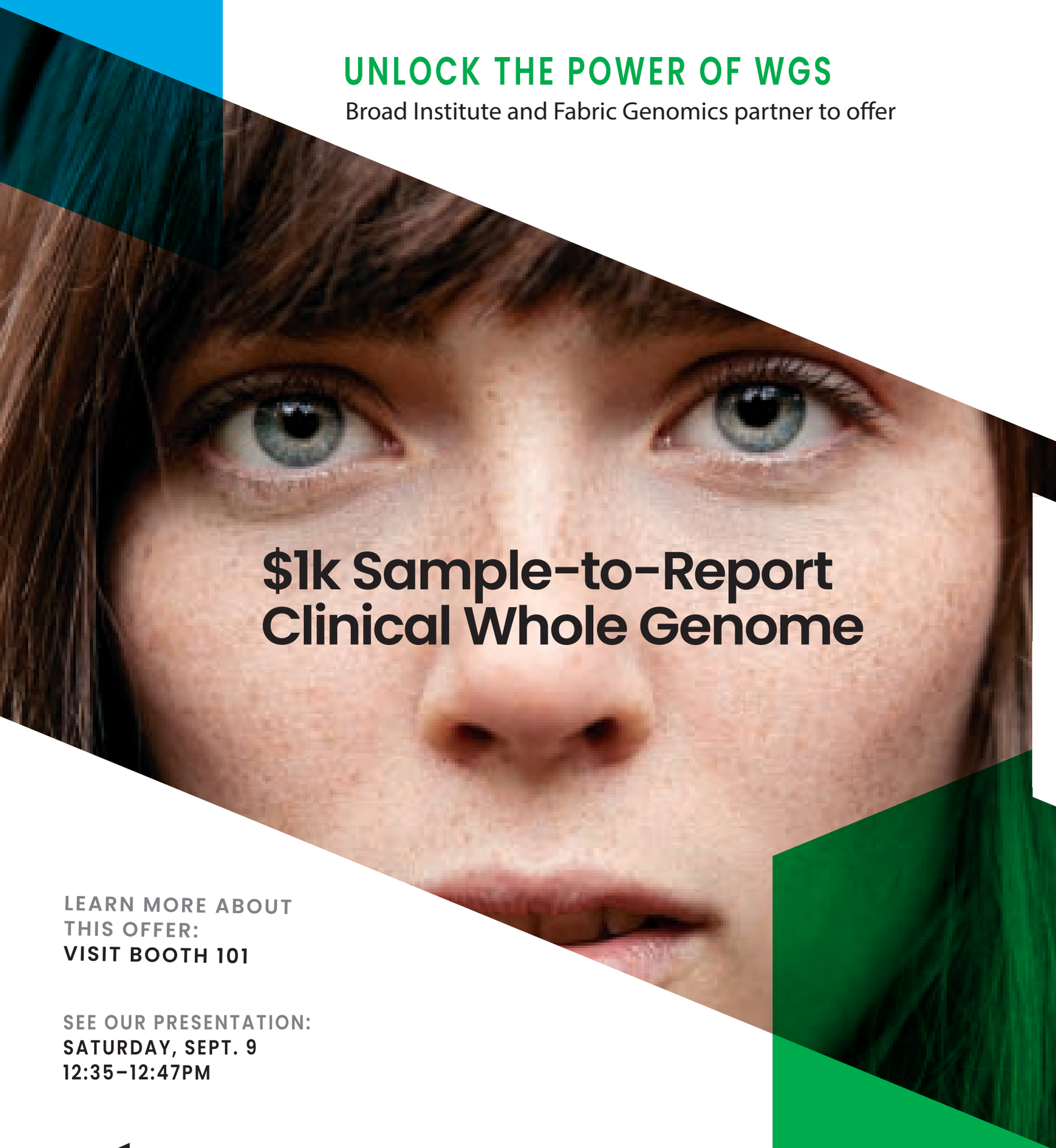
12:35 p.m. — 12:47 p.m.	**Ruby Sponsor – Fabric Genomics** **Michael Vishnevetsky, PhD, SVP of Business Development & Commercial Operations** "Unlock the power of clinical whole genome sequencing at a price that empowers adoption today"
12:47 p.m. — 12:59 p.m.	**Ruby Sponsor – Cellecta** **Alex Chenchik, PhD, President, Cellecta, Inc.** "TCR & BCR repertoire analysis and other approaches for the discovery of drug targets, resistance mechanisms and biomarkers"

Saturday, September 9, 2023 (Afternoon)

1:00 p.m. — 2:00 p.m.	**AGBT Lunch** Indigo Terrace
Session VII:	**Striving Towards Diversity and Equity in Precision Health (Chairs - Howard McLeod, Intermountain Healthcare and Catherine Cottrell, Institute for Genomic Medicine, Nationwide Children's Hospital)** Indigo Ballroom CDGH
2:00 p.m. — 2:30 p.m.	**Peter Hulick, Northshore University Health System** "Improving access to genetics through primary care"
2:30 p.m. — 3:00 p.m.	**Charles Rotimi, NIH Distinguished Investigator and Scientific Director, NHGRI** "Ancestry and disease in the age of precision medicine"
3:00 p.m. — 3:30 p.m.	**Claudia Gonzaga-Jauregui, International Laboratory for Human Genome Research, LIIGH, UNAM** "Genomic equity for health in Latin America"
3:30 p.m. — 3:50 p.m.	**Kathryn Gray, (Abstract Selected) University of Washington** "Multi-omics for precision medicine in preeclampsia in an admixed population"
3:50 p.m. — 4:00 p.m.	**Final Comments**

UNLOCK THE POWER OF WGS

Broad Institute and Fabric Genomics partner to offer



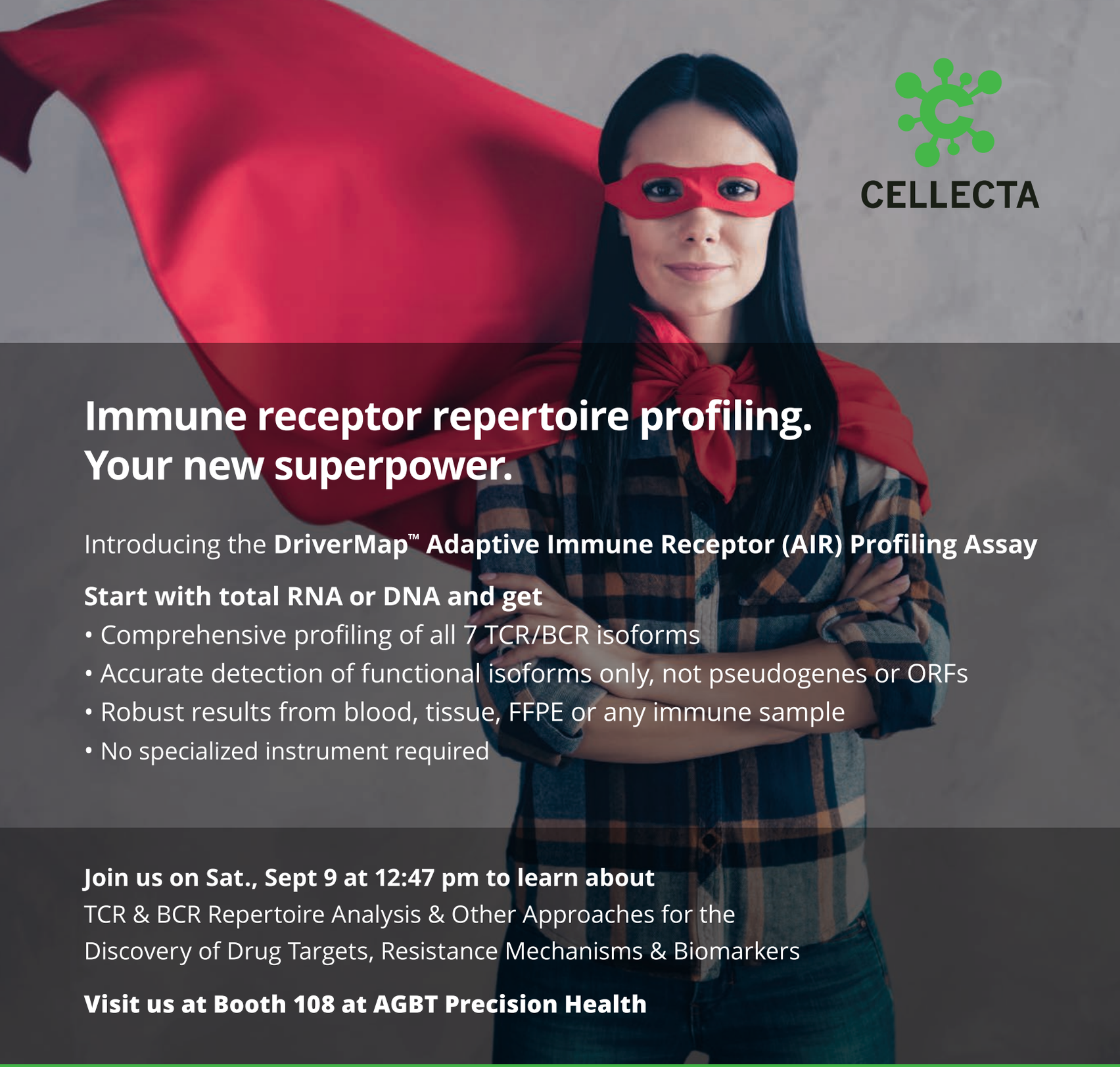
$1k Sample-to-Report Clinical Whole Genome

LEARN MORE ABOUT THIS OFFER:
VISIT BOOTH 101

SEE OUR PRESENTATION:
SATURDAY, SEPT. 9
12:35–12:47PM

FABRIC GENOMICS

BROAD INSTITUTE
CLINICAL RESEARCH SEQUENCING PLATFORM



CELLECTA

Immune receptor repertoire profiling. Your new superpower.

Introducing the **DriverMap™ Adaptive Immune Receptor (AIR) Profiling Assay**

Start with total RNA or DNA and get

- Comprehensive profiling of all 7 TCR/BCR isoforms
- Accurate detection of functional isoforms only, not pseudogenes or ORFs
- Robust results from blood, tissue, FFPE or any immune sample
- No specialized instrument required

Join us on Sat., Sept 9 at 12:47 pm to learn about
TCR & BCR Repertoire Analysis & Other Approaches for the Discovery of Drug Targets, Resistance Mechanisms & Biomarkers

Visit us at Booth 108 at AGBT Precision Health

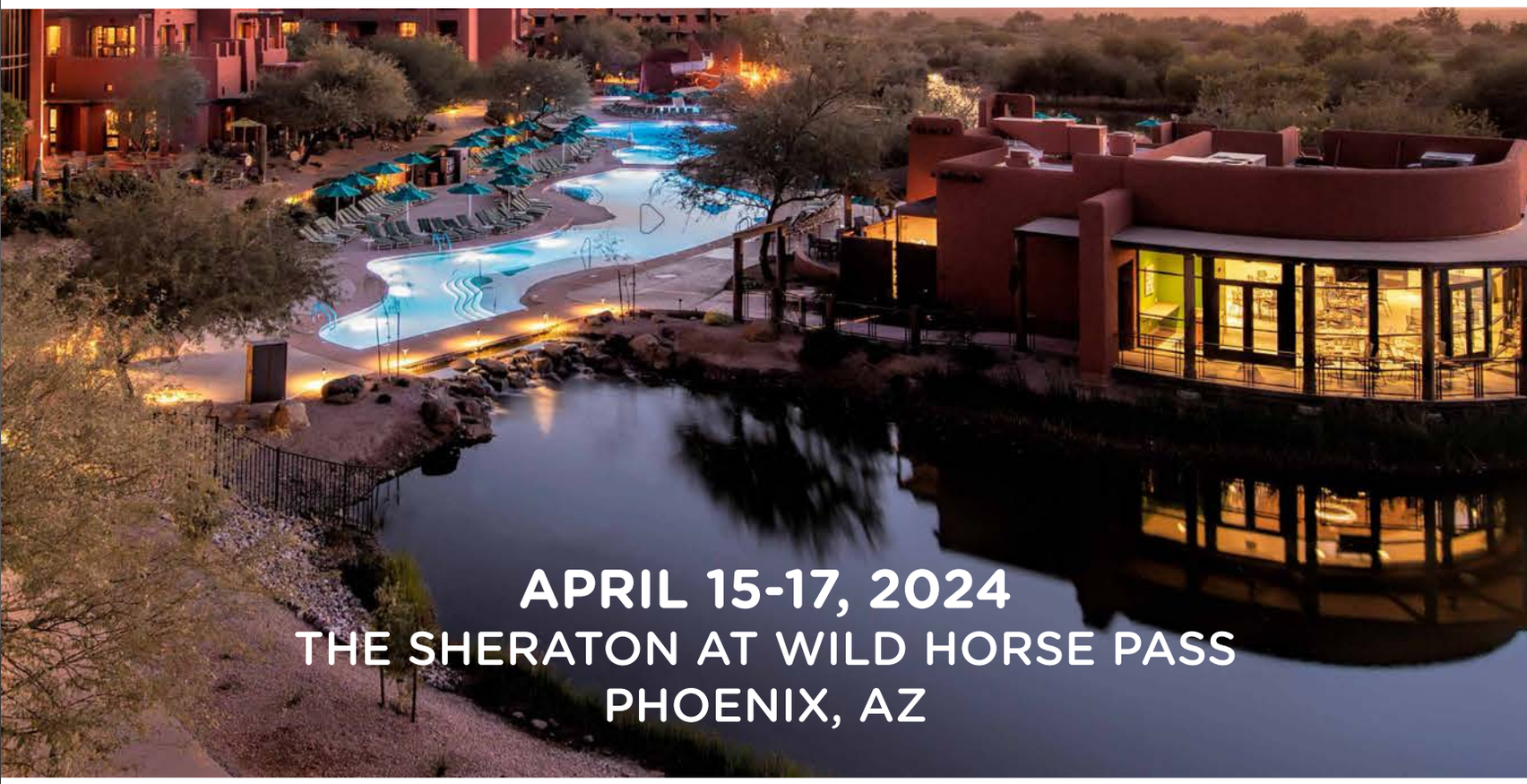
Who we are

Cellecta is a leading provider of genomic products and services. Our functional genomics portfolio includes gene knockout , knock-in, and knockdown screens, custom and genome-wide CRISPR and RNAi libraries, construct services, cell engineering, NGS kits as well as immune receptor repertoire and targeted expression profiling products and services.

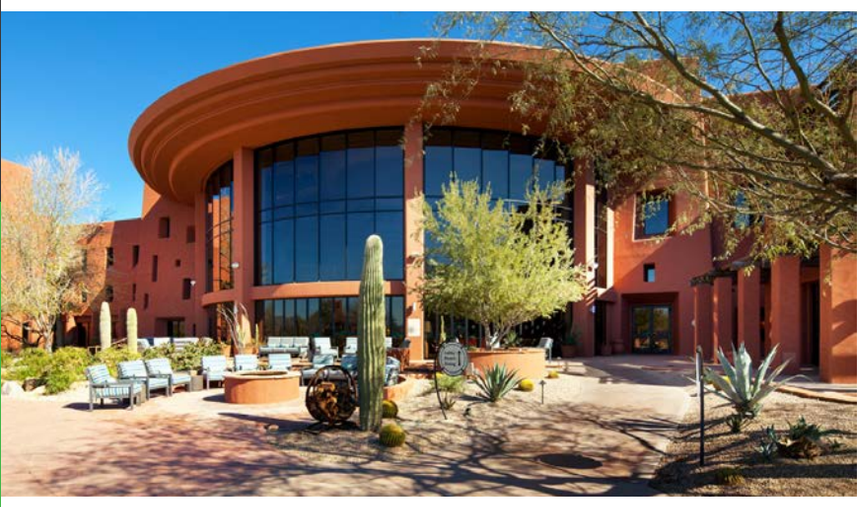
We help power your discovery efforts.
www.cellecta.com info@cellecta.com +1 650-938-3910

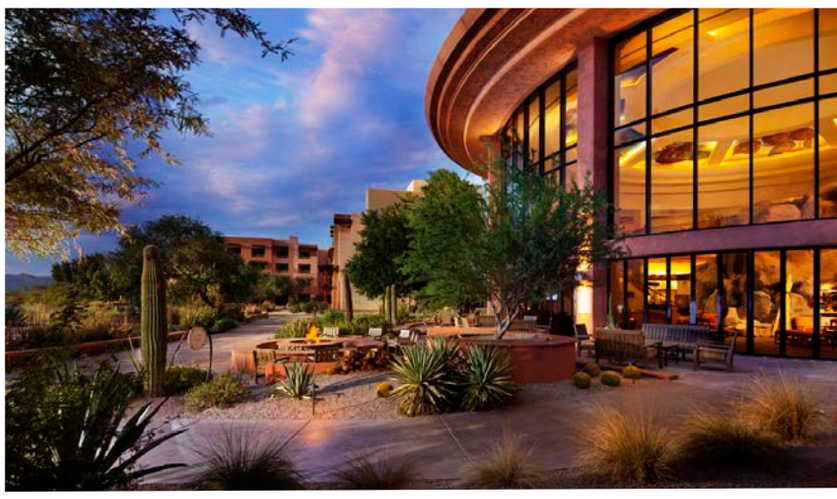
CELLECTA

AGBT2024™
AGRICULTURE



APRIL 15-17, 2024
THE SHERATON AT WILD HORSE PASS
PHOENIX, AZ





CALL FOR ABSTRACTS OPENS NOVEMBER 8TH

SEE OUR SPEAKERS AND LEARN MORE BY SCANNING
THE QR CODE, OR VISIT WWW.AGBT.ORG

A cloud-based, validated viral sequence analysis and database system to support antiviral drug development for viruses

POSTER 100

Lewyn Li[1], Casimir Saternos[1], Ran Yan[1], Joshua Cook[1] and Katie Kitrinos[1]

1 Assembly Biosciences, South San Francisco, CA, USA

Precision health (PH) aims to provide personalized healthcare solutions by integrating clinical data, genetic information, environmental factors, and lifestyle behaviors. To support ongoing clinical studies, we needed to develop an integrated database system to securely store and analyze patient viral sequences, while addressing a range of concerns including data privacy and security, standardization and quality control, and scalability within a flexible and cost-effective framework. Such data can help advance PH by providing results to enable treatment optimization, safety evaluation and regulatory approvals. However, the US FDA has published specific requirements on how Hepatitis B virus (HBV) sequencing data are to be reported.
We have developed "SeqDB", a custom-built analysis and database system of HBV DNA sequences to support our ongoing core inhibitor clinical development programs. Built entirely on the Amazon Web Service Cloud, SeqDB is a validated system that is compliant with FDA CFR 21 Part 11, EudraLex Annex 11 and other relevant predicate rules and regulations. Features of the current version include (1) the ability to upload Sanger sequences in a batch format with related clinical data via an easy-to-use graphical user interface; (2) automated sequence analysis to identify amino acid (aa) changes within a patient compared to pretreatment or reference sequences; (3) storage of original sequence files and detected aa changes information in a database; (4) the functionality to query aa changes from HBV DNA sequences collected from multiple clinical trials; (5) the generation of a summary report of aa changes observed in patients from a clinical trial. Modular by design, SeqDB is not a static system and new features will be developed and added as needed, including the ability to analyze sequences from additional viruses and to handle data from Next Generation Sequencing. With this cloud-based, validated system, sequence analysis, storage and reporting of HBV sequences from our core inhibitor clinical development programs have been unified under a single FDA/EMEA-compliant platform.

A novel suite of enzyme mixes enables robust hybrid capture sequencing from low quality FFPE DNA

POSTER 101

Pingfang Liu, Margaret R. Heider, Jian Sun, Brittany S. Sexton, Bradley W. Langhorst, Andrew Gray, Laura Blum, Lauren Higgins, Lixin Chen, Lynne M. Apone, Thomas C. Evans Jr., and Nicole M. Nichols

1 New England Biolabs Inc., Ipswich, MA 01938, USA

In cancer genomics, a common source of DNA is formalin-fixed, paraffin-embedded (FFPE) tissue from patient surgical samples, where in most cases high-quality fresh or frozen tissue samples are not available. FFPE DNA poses many notable challenges for preparing NGS libraries, including low input amounts and highly variable damage from fixation, storage, and extraction methods. Due to the high cost of sequencing and variability of coverage, regions of interest are often specifically enriched using hybrid capture-based approaches, but these methods require a high input of diverse, uniform DNA library to achieve the coverage required for somatic mutation identification in tumor samples.

We developed a new DNA repair enzyme mix, enzymatic fragmentation mix, and library amplification PCR master mix, optimizing the activities of these mixes using FFPE samples ranging from DIN 1.8 to 6.8 to maximize yield, WGS library quality, and target enrichment library performance. Combining DNA damage repair and a novel enzymatic fragmentation mix upstream of library preparation reduced the false positive rate in somatic variant detection by repairing damage-derived mutations, and also improved the library yield, quality metrics (including mapping, chimeras, and properly-paired reads), complexity, coverage uniformity, and hybrid capture library quality metrics. The new PCR master mix boosts the library yield without compromising library quality in FFPE-derived samples, allowing flexibility in the PCR cycles used to accommodate high-throughput processing of FFPE samples of highly varied quality. This library prep workflow was evaluated with multiple sequencing platforms including Illumina, MGI, and Element Biosciences.

This new suite of enzyme mixes allows even highly damaged FFPE samples to achieve high-quality libraries with sufficient input for hybrid capture. Increasing the useable reads and coverage enables robust detection of somatic variants as demonstrated using both reference standard DNA and patient-derived FFPE samples. Finally, the use of enzymatic fragmentation and a flexible PCR master mix make this FFPE library prep workflow compatible with high-throughput and automation-based workflows.

A pilot prospective clinical trial comparing BeginNGS, an artificial intelligence-enabled genome sequencing system for newborn screening of 410 childhood genetic diseases with rapid diagnostic genome sequencing

POSTER 102

Lauren Olsen[1,2], Katarzyna Ellsworth[1,2] Meredith Wright[1,2], Mei W. Baker[5], Sara Caylor[1,2], Thomas Defay[3], Annette Feigenbaum[1,2,4], Lucia Guidugli[1,2], Soofia Khan[1,2], Yong H. Kwon[1,2], Alisa Lam[1,2], Jennie Le[1,2], Jerica Lenberg[1,2], Lakshminarasimha Madhavrao[1,2], Rebecca Mardach[1,2,4], Danny Oh[1,2], Liana Protopsaltis[1,2], Muhammed Saad[1], Erica Sanford[1], Gunter Scharer[1], Jennifer Schleit[1,2], Brandan Schultz[1,2], Laurie D. Smith[1], Mari Tokita[1,2], Lucita Van Der Kraan[1,2], Kristen Wigby[1,2,4], Mary J. Willis[1], Mark Yandell[1,6], Charlotte A. Hobbs[1,2,4], Stephen F. Kingsmore[1,2]

1 Rady Children's Institute for Genomic Medicine, San Diego, CA 92123, USA.
2 Rady Children's Hospital, San Diego, CA 92123, USA.
3 Alexion, Astra Zeneca Rare Disease, Boston, MA 02210, USA.
4 Department of Pediatrics, University of California San Diego, San Diego, CA 92093, USA.
5 Wisconsin State Laboratory of Hygiene, Center for Human Genomics and Precision Medicine, Department of Pediatrics, University of Wisconsin School of Medicine and Public Health, Madison, WI 53706, US
6 Fabric Genomics, Inc., Oakland, CA 94612, USA.

Newborn screening (NBS) dramatically improves outcomes in selected, severe, childhood disorders by identification and treatment at or before symptom onset. Expansion of the NBS Recommended Uniform Screening Panel (RUSP) has lagged identification and approval of effective therapeutic interventions leading to delayed diagnosis and suboptimal outcomes in several hundred severe, early childhood onset, single locus (mendelian) genetic diseases. Begin Newborn Genomic Screening (BeginNGS) is one of several programs exploring the use of genome sequencing (GS) to supplement RUSP-NBS for these diseases. BeginNGS is unique with respect to the scope of NBS (currently 410 genetic diseases) and use of automated variant identification and virtual guidance regarding confirmatory testing, specialist referral, and urgent management. While clinical implementation is incomplete, here we report interim results of a first prospective clinical trial of BeginNGS. BeginNGS-1 is a scoping study to evaluate current performance and inform the design of a future, fully powered clinical study. Eligibility is limited to neonates aged 1 to 10 days who are admitted to the level IV NICU at Rady Children's Hospital, San Diego, and who are not suspected of having a genetic disease. Patients whose parents provide informed consent are enrolled. BeginNGS-1 enrollees receive RUSP-NBS and rapid, clinical diagnostic GS, which is performed immediately from dried blood spots with return of results to neonatologists or primary care pediatricians.

(continue to page 21)

(continued from page 20)

The BeginNGS screen will be performed after completion of enrollment by automated re-analysis of GS with a pre-qualified set of ~50,000 variants and a modified GEM tool (Fabric Genomics). Of 120 newborns screened thus far for eligibility, 83 (69%) were eligible for enrollment. 17 (20%) of 83 were discharged prior to enrollment, 32 (39%) of 83 declined enrollment, and 32 (39%) of 83 were enrolled. To date, diagnostic GS identified reportable findings in 6 (22%) of 27 newborns that may have clinical significance: two patients with likely pathogenic variants and four with variants of uncertain significance. Comparison of results of diagnostic GS, BeginNGS, RUSP-NBS, and confirmatory testing in the first 50 enrollees will be presented, together with implications for ongoing BeginNGS development and design of future clinical trials.

Advancing long-read nanopore genome assembly and accurate variant calling for rare disease detection

POSTER 103

Shloka Negi[1], Jean Monlong[1,2], Brandy McNulty[1], Emmanuèle Délot[3], Jonathan LoTempio[4], Seth Berger[3], Anne O'Donnell-Luria[5], Sara O'Rourke[1], Melanie O'Leary[5], Michael Talkowski[5], Heidi Rehm[5], Jonah Cool[6], Sara Simmonds[6], Eric Vilain[4], Karen H. Miga[1], Benedict Paten[1]

1 UC Santa Cruz Genomics Institute, University of California, Santa Cruz, 1156 High St, Santa Cruz, CA, USA
2 Institut de Recherche en Santé Digestive (IRSD), Université de Toulouse, INSERM, INRA, ENVT, UPS, Toulouse, France
3 Center for Genetic Medicine Research, Children's Research Institute, Children's National Hospital, Washington, DC, USA
4 Institute for Clinical and Translational Science, University of California, Irvine, CA, USA
5 Broad Center for Mendelian Genomics, Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA
6 Chan Zuckerberg Initiative, San Francisco, CA

The contribution of structural variants (SVs) to rare diseases is well-established, but their assembly and detection remain challenging with conventional short-read sequencing (SRS) approaches. To address this, we launched a pilot project utilizing ONT long-read sequencing (LRS) in a clinical research setting to identify rare genetic variants for rare disease diagnosis. We sequenced 98 samples (including 28 trios) using an efficient wet lab protocol, generating a high-quality long-read nanopore dataset. Leveraging the capabilities of LRS to generate haplotype-resolved genome assemblies and phased variant calls, our goal was to address cases where SRS lacked resolution due to genomic complexities and large structural variations. High-molecular-weight DNA extracted from stabilized blood of participants was subjected to LRS, targeting an N50 of 30kb and throughput of 100GB per promethION R10 flowcell. Sequencing data was then analyzed with Napu, a scalable computational pipeline we developed for the NIH-CARD project.

Napu performed reference-based mapping to detect small variants, generate haplotype-resolved methylation calls, and assembly-based SV identification. Finally, we developed a SnpEff-based, SV-aware clinical annotation pipeline to predict the functional impact of variants, narrowing down to a candidate set of clinically pathogenic rare variants for each proband. Phasing and haplotype-resolved assemblies enabled the identification of compound-heterozygous variants in some patients. A 6.7kbp SV and an SNV were found in compound-heterozygous orientation in the LHCGR gene, which was associated with Leydig cell hypoplasia, an autosomal recessive rare disorder.

(continue to page 23)

(continued from page 22)

While SRS couldn't resolve the gene/pseudogene conversions/30kbp deletions in the CYP21A2 gene causing Congenital Adrenal Hyperplasia, ONT-reads covered the region well, indicating the presence of 30kbp deletions in several patients. Additionally with LRS, alignment to the newly completed T2T Y-chromosome for the first time allowed disambiguation of X and Y pseudoautosomal regions where translocations occurred in some patients. Nanopore-based LRS outperformed SRS in the identification of large SVs, for example, in identifying rearrangements at the pseudoautosomal regions of the sex-chromosomes and establishing evidence for complex CYP21A2 gene/pseudogene rearrangements. Combined, these findings suggest that LRS technology holds promise for increasing the diagnostic yield of whole-genome sequencing for rare diseases.

Assessing adult neurogenesis activity and validating association to Alzheimer's disease

POSTER 104

Seungseok Kang[1], Saehoon Jung[2], Young-soo Chung[1], Yoo-jin Ha[1], Jeong Ho Lee[2], and Sangwoo Kim[1]

1 Department of Biomedical Systems Informatics, Brain Korea 21 Project, Yonsei University College of Medicine, Seoul, 03722, Republic of Korea
2 Graduate School of Medical Science and Engineering, Korea Advanced Institute of Science and Technology, Daejeon, 34141, Republic of Korea
These authors contributed equally.

Alzheimer's disease (AD) is a type of dementia caused by the excessive aggregation of beta-amyloid protein. Recent research suggests that genomic alterations may contribute to sporadic AD, which accounts for 95% of all cases. However, the direct involvement of somatic variants and their potential origin such as adult hippocampal neurogenesis (AHN) in AD brain remains unclear. In particular, decrease in AHN in AD patients has gained recognition in recent years, prompting extensive research. In this study, we employed laser capture microdissection to collect monoclonal-level specimens from the subgranular zone (SGZ) of the human dentate gyrus, aiming to isolate neuronal stem cells associated with AHN. We amplified low input DNA from the microdissected samples using the primary template-directed amplification (PTA) and performed whole-genome sequencing from 37 AD patients and 12 non-dementia individuals.

Although, no significant difference in mutation number between the AD and non-dementia group was observed, signature analysis revealed distinct mutational process. In the non-dementia group, we observed a significantly higher frequency ($p=0.029$) of the reactive oxygen species (ROS)-related signature SBS18 compared to the AD group that might arise due to sustained metabolism in normal neuronal stem cells. Furthermore, we found 8 AD-specific genes associated with neuron differentiation, axon degeneration, and the maintenance of excitatory synapses. In conclusion, our study provides evidence indicating impaired differentiation of neuronal stem cells related to AHN within the SGZ in AD patients.

Assessing the impact of batch correction using scanorama and Harmony on a neuroblastoma cell atlas

POSTER 105

Srikant Sarangi[1] , Javad Zahiri[1] , Nikolaos Patikas[2], Hongcheng Yao[2], Ilya Korsunsky[2] , Callen Bragdon[1], Martin Hemberg[2], Zachary Pitluk[1]

1 Paradigm4, 281 Winter Street, #360, Waltham MA 02451, USA
2 Brigham and Women's Hospital, 75 Francis Street, Boston MA 02115, USA

In recent years, reduced costs of single cell RNA sequencing (scRNA-Seq), has resulted in the generation of 10 to 100 patient datasets and tissue atlases. Even larger, population scale datasets will be necessary to derive accurate understanding of the interplay between RNA, DNA variation, protein abundance and metabolites. However, such datasets are compiled from multiple experiments that exhibit notable differences, including variations in platforms, technicians, and experimental conditions, among other factors. These differences result in technical batch effects that must be corrected to allow for meaningful biological analysis. To address this issue, numerous batch correction (BC) algorithms have been developed to remove batch effects while preserving the true biological variation between datasets.

In this study we compare two popular scRNA-Seq BC methods, scanorama and Harmony, and profile memory usage, computational complexity, and batch effect removal performance. The evaluation of the effect of batch correction was assessed using the k-nearest neighbor batch-effect test (kBET) and the average silhouette width (ASW). We used data from a Neuroblastoma Cell Atlas (DOI: 10.1126/sciadv.abd3311) which compared human neuroblastoma cells and a reference of normal human fetal medullary cells. Batch correction was applied across seven single cell datasets (Chromium 10x) taken from normal adrenal glands from five fetuses, and five neuroblastoma single cell datasets (Chromium 10x) obtained from five patients. We also investigated the removal of platform-dependent batch effects between the aforementioned 10x neuroblastoma datasets and single cell neuroblastoma data collected using CEL-seq2 from 16 patients. To enable this workflow, we implemented a pipeline using R and Python APIs through REVEAL, a bioinformatics platform developed by Paradigm4. The REVEAL platform uses an auto-scaling highly parallel cloud task execution engine (burst computing) and a distributed high-throughput file system (flexFS®), making it suitable for the processing of atlas scale scRNA-Seq datasets.

Our results provide deeper insights into the impact of batch correction using different algorithms, and rate limiting steps and limitations of the algorithms to support their refinement. Additionally, the study presents a general, scalable, and easily deployable workflow for assessing single-cell algorithms.

Association between tumor driver mutations and the presence of post-surgical circulating tumor DNA: Insights into tumor biology

POSTER 106

Noah Friedman[1], Sara L Bristow[2], Robert Burns[1], Minetta C. Liu[2], Alexey Aleshin[2], Helio Costa[1]

1 Natera, Inc., Data Insights, Austin, Texas, United States
2 Natera, Inc., Oncology, Austin, Texas, United States

While detection of molecular residual disease (MRD) using circulating tumor DNA (ctDNA) has been shown to be a prognostic biomarker for cancer monitoring, detection rates vary across tumor types and histology and are impacted by treatment. Ongoing clinical trials are investigating ctDNA for patient enrichment and as a surrogate endpoint for outcomes. Here we investigate the relationship between genomic alterations and MRD post-surgically. In total, 45,139 patients with solid tumors who underwent whole exome sequencing to design a personalized ctDNA assay (SignateraTM) through February 2023 were included. Fisher's Exact test with Bonferroni correction assessed associations between genomic alterations and anytime postsurgical ctDNA positivity over ctDNA-negativity. The most notable findings are presented here, with a more detailed landscape analysis available. We found 52 significant gene-cancer type associations (ctDNA risk: 23 increased; 29 decreased). TP53 accounted for the highest number of significant associations, with increased MRD rates in all cancer types except breast, lung, and skin. ctDNA rates were significantly different based on individual residues in 64 cases (26 increased; 38 decreased). Log-fold change in ctDNA positivity was higher for PIK3CA E545K in bladder (p=0.02), breast (p=.004), esophagogastric (p=0009) and lung (p=0.000003) cancers, while other PIK3CA residues in these cancer types were associated with no change or reduced rates of ctDNA positivity. In pancreatic cancer, ctDNA positivity rates with KRAS G12R (25%) were lower compared to KRAS G12V (37%, p=0.0003) and KRAS G12D (41%, p=0.0000005). Of note, pathognomonic driver mutations were more strongly associated with elevated risk of ctDNA detection in early-stage vs late-stage disease, such as KRAS in pancreatic cancer (early-stage: MRD-positive 72% vs MRD-negative 59%, p=.003; late-stage: MRD-positive 61% vs MRD-negative 61%, p=0.85) or TP53 in lung cancer (early-stage: MRD-positive 35% vs MRD-negative 26%, p<0.001; late-stage: 37% vs 36%, p=0.65). Our study presents one of the largest datasets to date with detailed landscape analysis demonstrating that genomic drivers in tumors can influence ctDNA detection rates throughout disease treatment and surveillance. Early vs. late-stage dependent differences in the impact of genomic alterations on ctDNA detection rates may provide biomarkers for enhanced clinical risk staging in addition to histopathological factors.

Automated CUT&RUN brings scalable epigenomic profiling to personalized medicine

POSTER 107

Keith E. Maier[1] , Matthew R. Marunde[1] , Vishnu U. Sunitha Kumary[1], Carolina P. Lin[1], Danielle N. Maryanski[1], Liz Albertorio-Saez[1], Dughan J. Ahimovic[2] , Michael Bale[2], Juliana J. Lee[3], Bryan J. Venters[1], Michael-Christopher Keogh[1], Immunological Genome Consortium

1 EpiCypher Inc., Durham, NC, USA
2 Weill Cornell Medicine, Department of Pathology and Laboratory Medicine, New York, NY, USA
3 Harvard Medical School, Department of Immunology, Boston, MA, USA

Understanding chromatin dynamics is central to to the advancement of precision medicine. Genome-wide association studies (GWAS) and corresponding transcriptomics research have been leveraged to identify and classify some disease risk variants. However, >90% of GWAS-identified variants are found in non-coding regions of the genome, indicating they likely affect the regulatory machinery that governs chromatin structure. To date, efforts to understand the effects of non-coding variants on chromatin dynamics have been focused on chromatin accessibility (ATAC-seq) and DNA Methylation profiling. However, these assays provide a largely binary view of chromatin (open/closed) and often fail to provide mechanistic insight into disease etiology. Epigenomic features – such as histone post-translational modifications (PTMs) and chromatin-associated proteins – mark distinct genomic compartments (e.g., promoters, enhancers) and regulate chromatin structure, gene expression, and cell function. Mapping these features provides a rich context to study cell fate and has great potential for discovering new biomarkers and drug targets. Despite this, efforts to integrate epigenomics into precision medicine research have been hampered by the poor sensitivity, high background, and low throughput of traditional chromatin mapping technology (i.e., ChIP-seq).

Here, we present autoCUT&RUN, a high throughput assay for rapid, ultra-sensitive profiling of epigenomic features from FACS-isolated primary cells or tissues. This workflow generates reliable profiles from <10,000 cells per reaction, and is supported by a rigorous optimization strategy, high-quality antibodies, and quantitative spike-in controls. As part of a multi-site collaboration with the Immunological Genome Consortium, we used our autoCUT&RUN platform to build a comprehensive epigenomic database of mouse immune cells - composed of >1,500 epigenomic profiles from >100 different FACS-isolated primary immune cell types. These studies set the stage to leverage high-throughput epigenomics for multi-site clinical studies to identify and/or characterize novel biomarkers for precision medicine applications.
Further, we will discuss our integration of CUT&RUN and Enzymatic Methylation-sequencing (CUT&RUN-EM) to deliver a true multiomic view of the co-occurrence of epigenomic features and DNA methylation, which can provide novel insights into cellular mechanisms and sample heterogeneity.

Automated Prioritization of Sick Newborns for Rapid Whole Genome Sequencing Expedites Diagnosis

POSTER 108

Edwin Juarez[1], Bennett Peterson[2], Sheldon Gilmer[3], Shauna Briscoe[3], Curtis Beebe[3], Rebecca Reimers[1,3], Erica Sanford-Kobayashi[1], Chris Tackaberry[4], Mark Yandell[2], and Charlotte Hobbs[1]

1 Rady Children's Institute for Genomic Medicine, San Diego, CA, USA
2 University of Utah, Department of Human Genetics, Salt Lake City, UT, USA
3 Rady Children's Hospital, San Diego, CA, USA 4Clinithink, London, England

Genetic disorders are a major cause of death and disability for infants admitted to the Neonatal Intensive Care Unit (NICU). Rapid Whole Genome Sequencing (rWGS) can help swiftly diagnose patients with genetic disease and facilitate care tailored to each patient. Deciding whom to sequence is challenging.

This study deployed a clinical decision support tool called the Mendelian Phenotype Search Engine (MPSE) at Rady Children's NICU. The MPSE, developed by University of Utah, utilizes a Natural Language Processing (NLP) solution provided by Clinithink, and Machine Learning (ML) to compute a patient-specific score summarizing how similar their phenotype is to other patients who have previously received rWGS.

This study consisted of two phases: In phase 1 critically ill infants were nominated for rWGS as per current best practice. In phase 2, attending physicians were provided with a daily report containing MPSE scores for each NICU patient, and nominations were accepted based on the same criteria as phase 1. The sole difference between phase 1 and phase 2 was that the MPSE Scores were made available to attending physicians in phase 2.

Phase 1 lasted 14 weeks and Phase 2 lasted 38 weeks. Throughout the study, there were 118 nominations: 27 in phase 1 (1.93 nominations/week) and 91 in phase 2 (2.39 nominations/week). From these nominations, 98 patients were enrolled and underwent rWGS; 25 from phase 1 (1.79 enrollments per week) and 73 from phase 2 (1.92 enrollments per week). In total 29 enrolled infants had pathogenic or likely pathogenic variants consistent with their phenotype, with 6 occurrences in phase 1 (diagnostic yield of 6/25 = 0.24) and 23 in phase 2 (diagnostic yield of 23/73 = 0.32). Most notably, the average time for patients to be referred for sequencing was reduced by almost half (114 hours in Phase 1; 71 hours in Phase 2). Thus, our clinical deployment of MPSE helped to reduce the time to diagnosis of critically ill infants by approximately two days.

These findings underscore the immediate impact that NLP and machine learning can have in assisting NICU medical teams with their clinical decision-making processes.

Automated Processing of FFPE Samples for Single Nuclei RNA-Seq

POSTER 109

Stevan Jovanovich[1], Nate Pereira[1], Catherine Snopkowski[2], Danielle Meyer[1], Brigita Meškauskaitė[2], Meril Takizawa[2], Zaki Abou-Mrad[3], Kenny K H Yu[3], Viviane Tabar[3], Ignas Masilionis[2], Nabiha Khan[1], John Bashkin[1], and Ronan Chaligne[2]

1 S2 Genomics, Inc., Livermore, CA, USA
2 Single-cell Analytics Innovation Lab Memorial Sloan Kettering Cancer Center, NY, USA
3 Department of Neurosurgery, Memorial Sloan Kettering Cancer Center, NY, USA

S2 Genomics has developed the SingulatorTM 100 and 200 Systems to automate the isolation of singulated cells or nuclei from fresh, frozen, or OCT samples using disposable cartridges, reagents, and customizable protocols. The systems have been applied to prepare cells or nuclei from a wide range of human tissues (healthy, tumor, organoid) and wide range of organisms including mouse, rat, chicken, pigs, insects, snails, zebrafish, planaria, and plants.

We have now extended these applications to the automated processing of FFPE slices into singulated nuclei on prototype Singulator systems. The systems perform all deparaffinization and rehydration steps, optional crosslink reversal, followed by isolation of nuclei. The nuclei are processed using a commercially available probe-based kit (Chromium Single Cell Gene Expression Flex, 10X Genomics) to create single nuclei gene expression libraries. Sequencing of a brain library made from FFPE showed equivalent results of automated deparaffinization and rehydration and manual processing.

Data will be presented on the preparation and analysis of snRNA-Seq libraries from healthy brain and different human tumor clinical samples including colorectal carcinomas and glioblastomas prepared from FFPE slices.

Benchmarking the performance of tumor-only somatic variant detection algorithms with the G4 sequencing platform

POSTER 110

Timothy Looney

Background
Next-generation sequencing (NGS) based identification of actionable somatic variants in cancer samples is a critical component of precision oncology and immunotherapy, though also one of the most technically challenging applications. To date, a variety of open-source algorithms have been developed for the detection of somatic variants in tumor-only samples, each with potential strengths. Here we benchmark the performance of tumor-only variant calling algorithms with the G4 sequencing platform using a large somatic variant reference sequencing dataset.

Methods
Control gDNA consisting of an equimolar pool of 23 reference cell lines derived from the 1000 genomes project was used for library preparation via the IDT xGen cfDNA kit, followed by enrichment via the IDT Pan Cancer Panel (800kb), then sequenced to ~20,000x mean coverage via the G4 F2 flow cell (150M reads/flow cell). Single read family correction was performed using the fgbio suite of tools, then corrected reads were analyzed using algorithms leveraging heuristics (Vardict, VarScan2), haplotyping with local assembly (Freebayes, Mutect2), and site-specific error (SSE) estimation (Outlyzer). An ensemble machine learning approach (SomaticSeq) using these individual callers was also used. Separately, raw reads were processed using the UMI-aware end-to-end pipeline MAGERI. Performance was measured via F1 score as determined by som.py.

Results
Heuristic-based algorithms outperformed other approaches with respect to precision and accuracy, led by Varscan2. Integrated UMI-analysis workflows underperformed in comparison with those employing UMI-corrected reads. Performance of each respective tool was in line with historical performance with Illumina data. An ensemble machine learning variant calling approach leveraging heuristic, SSE and haplotyping resulted in the highest overall F1-score of 0.86 for SNPs and Indels combined.

Conclusions
We have extensively surveyed the tumor-only somatic variant algorithm space, identifying the best performing individual and combination algorithms for somatic variant detection on the G4, and confirming that the G4 delivers variant calling results in line with the current state-of-the art. We anticipate that the rapid turnaround, flexibility and high-accuracy of the G4 will significantly reduce turnaround for the most time-sensitive NGS applications.

Benefits of integrating an open-source knowledge-base in a precision oncology workflow

POSTER 111

Cameron J. Grisdale[1] , Erin Pleasance[1] , Connor Frey[2], Caralyn Reisle[1], Laura M. Williamson[1], Veronika Csizmok[1], Kathleen Wee[1], Yaoqing Shen[1], Melika Bonakdar[1], Greg Taylor[1], Kilannin Krysiak[3], Jason Saliba[4], Arpad M. Danos[4] , Adam C. Coffman[4], Susanna Kiwala[4], Joshua F. McMichael[4], Janessa Laskin[2], Malachi Griffith[4], Obi L. Griffith[4], Steven J.M. Jones[1]

1 Canada's Michael Smith Genome Sciences Centre, Vancouver, BC, Canada
2 BC Cancer, Department of Medical Oncology, Vancouver, BC, Canada
3 Washington University School of Medicine, Department of Pathology and Immunology, St. Louis, MO, USA
4 Washington University School of Medicine, Department of Medicine, St. Louis, MO, USA

Advancements in next-generation sequencing (NGS) technologies have brought about a paradigm shift in precision oncology, yet the extraction and interpretation of clinically relevant alterations from these extensive NGS datasets pose a considerable challenge that needs to be addressed. The Personalized OncoGenomics (POG) program at BC Cancer utilizes whole genome and transcriptome analysis (WGTA), providing a comprehensive view of the molecular biology of advanced cancer patient tumours, with over 1000 patients enrolled to-date. This analysis relies on curated clinical knowledgebases linking cancer variants and their clinical relevance, but the breadth and utility of these can be limited by access restrictions or missing information. CIViC (Clinical Interpretation of Variants in Cancer; civicdb.org) is an open-access, expert moderated, crowd-sourced knowledgebase of clinically relevant cancer variants that aims to address these limitations and is one of several sources used for variant interpretation in POG. Based on a retrospective cohort of POG cases, we evaluated the knowledgebase coverage of genes and variants involved in treatment recommendations from the molecular tumour board (MTB) as well as those suggested by genome analysts. We found more than 95% of patients had an alteration considered clinically actionable by the MTB, demonstrating the benefit of WGTA paired with open-source automated variant matching and reporting software. Clinical interpretations derived from CIViC represented nearly 50% of therapeutic evidence reported at the MTB, emphasizing the role of open-access knowledge in precision oncology. Additionally, by prospectively examining the levels and sources of evidence in patient reports, we identified areas where additional knowledgebase curation would be most clinically impactful, highlighting genome signatures as a critical area for future curation efforts. Finally, we considered the impact of quality of evidence on MTB recommendations and patient treatments.

ABSTRACT SELECTED TALK

Cardiovascular disease biomarkers derived from circulating cell-free DNA methylation

Rafael R. C. Cuadrat[1], Adelheid Kratzer[2,3], Hector Giral Arnal[2,3], Anja C. Rathgeber[1], Katarzyna Wreczycka[1], Alexander Blume[1], Irem B. Gündüz[1], Veronika Ebenal[1], Tiina Mauno[1], Brendan Osberg[1,4], Minoo Moobed[2,3], Johannes Hartung[2,3], Kai Jakobs[2,3], Claudio Seppelt[2,3], Denitsa Meteva[2,3], Arash Haghikia[2,3,4], David M. Leistner[2,3], Ulf Landmesser[2,3,4], Altuna Akalin[1]

1 Bioinformatics & Omics Data Science Platform, Berlin Institute for Medical Systems Biology, Max-Delbrück-Center for Molecular Medicine Berlin, Berlin, Germany
2 Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Department of Cardiology, Campus Benjamin Franklin, Berlin, Germany;
3 DZHK (German Centre for Cardiovascular Research), partner site Berlin
4 Berlin Institute of Health (BIH), Berlin, Germany
5 Current address: Max-Planck-Institut für molekulare Genetik, Berlin, Germany

Acute coronary syndrome (ACS) represents a significant global health burden, contributing to substantial morbidity and mortality rates. Prompt identification and appropriate management of ACS are crucial for improving patient outcomes. In this study, we explored the utility of circulating cell-free DNA (ccfDNA) methylation profiling as a novel approach for differentiating between the types of ACS and identifying potential biomarkers.

The differentiation of ACS subtypes, including non-ST-elevation myocardial infarction, ST-elevation myocardial infarction, and unstable angina, is essential for tailoring precise treatment strategies. Traditionally, clinical findings, such as electrocardiograms and plasma biomarkers, have been used to classify ACS types. However, we propose ccfDNA as a more accurate marker. Through comprehensive analysis, we discovered hundreds of methylation markers associated with ACS subtypes, validated in an independent cohort. Notably, these markers were frequently linked to genes involved in cardiovascular conditions and inflammation.

Our findings demonstrate the potential of ccfDNA methylation profiling as a non-invasive diagnostic tool for acute coronary events. The use of methylation-based biomarkers offers several advantages, including accessibility, scalability, and the ability to repeat similar analyses for various diseases. Furthermore, we extended this methodology to heart failure and obstructive coronary artery disease, yielding similar promising results.

(continue to page 33)

(continued from page 32)

In conclusion, this study highlights the promise of ccfDNA methylation profiling as a valuable tool for multiple heart diseases and identifying associated biomarkers. By leveraging this methodology, we can enhance diagnostic accuracy, facilitate personalized treatment selection, and advance our understanding of the underlying mechanisms driving cardiovascular diseases. The non-invasive nature of this approach opens new avenues for early detection and monitoring of acute and chronic cardiovascular conditions, ultimately leading to improved patient care and outcomes.

CLEMENT: Genomic decomposition and reconstruction of non-tumor subclones

POSTER 112

Young-soo Chung[1] , **Seungseok Kang** [1]**and Sangwoo Kim**[1]

1 Department of Biomedical Systems Informatics, Brain Korea 21 PLUS Project for Medical Science, Yonsei University College of Medicine, Seoul, 03722, Republic of Korea

Genome-level clonal decomposition of a single specimen has been widely studied; however, it is mostly limited to cancer research. In contrast, the decomposition of non-tumor samples is challenging due to the following reasons: i) the variant allele frequency (VAF) of mutations is low, resulting in numerous false positives and false negatives, ii) there is significant genomic similarity between the clones, and iii) the absence of copy number alteration features. Recently, interest has arisen in non-cancer microscopic tissues, such as tissue-level mosaicism or developmental tracing.

In this study, we developed a new algorithm called CLEMENT, which accurately decomposes and reconstructs multiple subclones in genome sequencing of non-tumor (normal) samples. CLEMENT utilizes the Expectation-Maximization (EM) algorithm with optimization strategies specifically designed for non-tumor subclones. These strategies include false variant call identification, non-disparate clone fuzzy clustering, and clonal allele fraction confinement. In simulations and in vitro cell line mixture data, CLEMENT outperformed current cancer decomposition algorithms in estimating the number of clones (root-mean-square-error = 0.39–1.56 vs. 1.36–3.72) and in the variant-clone membership agreement (82.5% vs. 73.0–76.7%).

Additional testing on human multi-clonal normal tissue sequencing confirmed the accurate identification of subclones originating from different cell types. Applying CLEMENT to the adrenal gland (AG) revealed clonal development along the different layers of the AG and confirmed the presence of stem cell clusters in Zona Fasciculata. We expect that CLEMENT will serve as a crucial tool in the currently emerging field of non-tumor genome analysis.

Collaborative, longitudinal genomic diagnostic care in large cohort projects

POSTER 200

Alistair Ward, **Isabelle Cooperstein, Tony Di Sera, Stephanie Georges, Anders Pitman, Marti Tristani-Firouzi, Gabor Marth**

1 Canada's Michael Smith Genome Sciences Centre, Vancouver, BC, Canada
2 BC Cancer, Department of Medical Oncology, Vancouver, BC, Canada
3 Washington University School of Medicine, Department of Pathology and Immunology, St. Louis, MO,
USA
4 Washington University School of Medicine, Department of Medicine, St. Louis, MO, USA

Patients presenting with complex phenotypes represent some of the most challenging diagnostic cases. Evolving phenotypes, bioinformatic and biological resources demand these cases are regularly monitored to ensure diagnosis can be achieved at the earliest opportunity. Further, diagnosis requires the combined expertise of large teams comprising treating physicians, genetic counselors, medical geneticists, diagnostic pathologists, bioinformaticians and variant analysts that bring together knowledge of the patient's family history and clinical presentation; etiology of genetic disease; computational analysis; and variant interpretation to expertly adjudicate the role of variants for unique patients. Existing software tools typically cater to those with computational expertise and can exclude other critical team members from contributing their full expertise to the diagnostic process.

We are developing Calypso; a comprehensive software system to provide a state-of-the-art platform to support diagnostic teams in a collaborative web-based environment. Combining a computational backend for analysis tasks; an easy-to-use web-based front-end integrating popular iobio tools; and a communication interface, Calypso will help team-based, longitudinal genome diagnostics become a reality. Longitudinal analysis will include variant re-analysis, and re-annotation pipelines that ensure changes in a patient's phenotypes, or underlying genetic knowledge (e.g. updates to databases like ClinVar) are made available to the diagnostic team as early as possible.

Cohort dashboards allow case information in large projects (for example the Undiagnosed Diseases Network, UDN) to be visually interrogated, and variants present in any cases in the cohort to be interrogated.

(continue to page 36)

(continued from page 35)

Case dashboards provide summaries of patient cases, including a visual timeline of a patient case and notifications of outstanding actions. This will include lists of required tasks including reviewing variants that have been flagged by the re-analysis / re-annotation pipelines.

Integrated analysis tools include the popular iobio apps, IGV, and a tool to identify phenotypically "similar" patients in the cohort will help all team members leverage all case and cohort data to reach diagnosis for every patient.

Calypso has been deployed at the University of Utah in our rapid sequencing program and, as part of UDN Phase 3, will provide a network level platform to support a high level of collaboration across sites.

CoLoRSdb: An integrated variation frequency landscape in long-read human genomes

POSTER 201

Qiuhui Li[1] on behalf of the Consortium of Long Read Sequencing (CoLoRS)

1 Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA

In the past two decades, human genetics studies have unraveled tremendous variation within the human genome spanning from single nucleotide variations (SNVs) to complex structural rearrangements. Numerous studies, including gnomAD, ClinVar and GTEx have deepened our understanding of variants in the human genome and elucidated their roles in gene regulation, ethnic diversity and genetic diseases. Comprehensive profiling of genetic variation is critical to unlocking the complexity of the human genome and facilitating progress in genomic research. However, investigating complex structural variations and certain SNVs - particularly those located in highly repetitive regions - remains a challenge due to short-read sequencing constraints. Addressing this need, we introduce the Consortium of Long Read Sequencing (CoLoRS; colorsdb.org), an open coalition of international researchers focused on cataloging and providing frequency information for all classes of variation using long-read whole genome sequencing. Long-read sequencing produces reads tens of kilobases long that can span difficult-to-assemble regions and improve variation detection, especially structural variations and other complex variations. Long-reads also have improved power to access variation within homologous sequences and highly repetitive regions that are out of reach for short reads. The genomic data to populate the database is extracted from pre-existing and on-going research projects conducted by members of the consortium. Sequencing reads are then processed, via standardized pipelines at the individual sites or within the AnVIL cloud platform, to summary level data for public release.

Here, we present our initial work cataloging variations in >1,000 long-read sequenced human genomes aggregated from several studies. The WDL-based processing pipeline includes DeepVariant for SNVs and small indel detection, pbsv & Sniffles for structural variant identification, TRGT for tandem repeat analysis, and HiFiCNV for copy number detection. After variant detection, we plan to assess the impact of cataloged variants with eQTL analyses, especially within medically relevant genes. We also use the database to filter candidate variants from on-going long-read clinical studies. We expect CoLoRSdb will serve as a critical resource for the global scientific and clinical research community, advancing the development of precision medicine and promoting the integration of genomics into clinical practice.

Complete genomes of a three-generational pedigree to expand studies of genetic and epigenetic inheritance of centromeres

POSTER 202

Monika Cechova[1], Sergey Koren[2], Julian K. Lucas[1], Rebecca Serra Mari[3], Mobin Asri[1], David Porubsky[4], Jordan M. Eizenga[1], Brandy McNulty[1], Andrey Bzikadze[5], Shloka Negi[1], Christopher Markovic[6], Tamara Potapova[7], Jennifer L. Gerton[7], Pavel A. Pevzner[8], Evan E. Eichler[4], Benedict Paten[1], Adam M. Phillippy[2], Ting Wang[9], Nathan O. Stitziel[10] , Robert S. Fulton[6], Tobias Marschall[3], Karen H. Miga[1]

1 UC Santa Cruz Genomics Institute, University of California, Santa Cruz, 1156 High St, Santa Cruz, CA, USA
2 Genome Informatics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA
3 Institute for Medical Biometry and Bioinformatics, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany
4 Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195, USA
5 Graduate Program in Bioinformatics and Systems Biology, University of California San Diego, San Diego, CA, USA 6 McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO 63108, USA
7 Stowers Institute for Medical Research, Kansas City, MO, USA
8 Department of Computer Science and Engineering, University of California San Diego, San Diego, CA, USA
9 Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110, USA
10 Department of Medicine, Washington University School of Medicine, St. Louis, MO 63110, USA

The complete genomes provide an opportunity for a new biological discovery in repetitive parts of the genomes that were frequently missing, incomplete, or misrepresented in the past, such as centromeres. Here, we present a study aiming to release complete, T2T diploid assemblies of four individuals, representing a three-generational pedigree of African-American ancestry. To this end, we have utilized a combination of high-coverage, long-read (51x HiFi, PacBio, and 71x Nanopore Ultra-long data 100kb+, Oxford Nanopore Technologies) and short-read paired technologies for haplotype phasing (including 76x coverage Omni-C, Cantata Genomics, 39x paired read parental data for trio-based phasing, as well as Strandseq and Pore-C data). Using two iterative, graph-based methods with ultra-long nanopore integration (verkko (Rautiainen et al. 2023) and hifiasm-UL (Cheng et al. 2022)), we achieved a high number (e.g. 28/46) of automated telomere-to-telomere (T2T) chromosome assemblies, with additional improvements with the two assembly methods combined. This allowed us to study transgenerational inheritance in biologically critical regions that are highly repetitive and present copy number variation within the population. We provide evidence for genetic and epigenetic inheritance across large tandem repeats that define human centromeres.

(continued on page 39)

(continued from page 38)

For example, we found that the centromeric sequence for chromosome 12 was 100% identical across the three generations, spanning the length of 3,321,121 basepairs, allowing us to study the biological variation in methylation patterns in this region, such as the variation in CDR (Centromere Dip Region), marked by large dips in methylation that underlie the binding of the centromeric protein CENP-A. Moreover, we found preliminary evidence of the enrichment of another modification, 5-Hydroxymethylcytosine (5hmC), in the peri(centromeric)

region, and especially in the flanking HSat3 satellite arrays. Additionally, our assemblies captured several rDNA arrays, such as on chromosomes 21 and 22. In summary, these results provide a high-quality multi-generational pedigree that serves as a community resource for tracing of transgenerational inheritance, as well as genetic and epigenetic variation of centromeres, satellite DNA, and rDNA arrays.

Comprehensive characterization of FMR1 mutation status with combination short-read and nanopore long-read confirmation

POSTER 203

Jiayi Sun[1], Muh-ching Yee[1], Stephanie Greer[1], Donna Hongo[1], Brynn Levy[1,2], Premal Shah[1], Matthew Rabinowitz[1,3], Akash Kumar[1], Kate Im[1]

1 MyOme, Inc., Menlo Park, CA, USA
2 Columbia University Irving Medical Center, Department of Pathology and Cell Biology, New York, NY, USA
3 Natera, Inc. San Carlos, CA, USA

Fragile X Syndrome (FXS), the most commonly inherited X-linked neurodevelopmental disorder, results from the expansion of CGG repeats within the 5' UTR of the fragile X mental retardation 1 (FMR1) gene triggering methylation and transcriptional silencing. Our study presents a novel pipeline for assessing FMR1 status, combining short-read (SR) and long-read (LR) sequencing techniques to determine CGG repeats and methylation status. Illumina 30x PCR-free HiSeqX sequencing of 7 control samples and 14 case samples (1 gray, 45-54 repeats; 7 premutation, 55-199 repeats; 6 full-mutation, >200 repeats) from Coriell were used for testing our SR approach which detected potentially affected versus normal samples. Oxford Nanopore (ONT) LR sequencing was performed on a set of 2 normal, 2 premutation, 2 full mutation samples. LR barcoded libraries were prepared by tagmentation of genomic DNA to an average size of 15 kb. Sequencing was done using Adaptive Sampling targeting the FMR1 gene. STR repeat detection in LR data was performed using Tandem Genotypes. Methylation status of the FMR1 promoter region was visualized using the Integrative Genomics Viewer (IGV). Our SR approach correctly identified affected samples, but did not accurately assess the number of CGG repeats. Using LR sequencing, we were able to correctly identify the number of alleles for normal samples within 1 repeat. Pre-mutation samples, one with 70/23 repeats and the other with 56 repeats, were identified as 78/23 and 55-57 respectively. For full mutation alleles (>600 repeats), the two samples were within 10% and 15% of the Coriell reported number of repeats. We also confirmed methylation matched expansion status of alleles. We demonstrated the utility of combining SR and LR sequencing to accurately identify repeat number and methylation status in the context of FMR1. Underlying this study is a generalized approach towards confirmation of STRs and other variants in difficult-to-detect regions, excluding the need for costly and lengthy methods of orthogonal confirmation. As LR sequencing technology evolves, we envision a future in which it is used as the single platform for diagnosis and confirmation of genetic disease, potentially shortening the time to a medically actionable diagnosis.

Comprehensive characterization of second-hit variants in vascular anomalies and neurocutaneous disorders using paired exome analysis

POSTER 204

Ying-Chen Claire Hou[1,2], Bhuvana A. Setty[3,4], Anna P. Lillis[5,6], Ibrahim Khansa[7], Gregory D. Pearson[7], Esteban Fernandez Faith[4,8], Archana Shenoy[2,9], Sonja Chen[2,9], Brandon S. Stone[10], Erica Macke[1], Gregory Wheeler[1], Benjamin Kelly[1], Mari Mori[4,10], Mariam Mathew[1,2,4], Kathleen M. Schieffer[1,2,4], Elizabeth Varga[1], Elaine R. Mardis[1,4], Catherine E. Cottrell[1,2,4]

1 The Steve and Cindy Rasmussen Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, OH, USA
2 Department of Pathology, The Ohio State University, Columbus, OH, USA
3 Division of Hematology/Oncology/BMT, Nationwide Children's Hospital Columbus, OH, USA
4 Department of Pediatrics, The Ohio State University, Columbus, OH, USA
5 Department of Radiology, Nationwide Children's Hospital Columbus, OH, USA
6 Department of Radiology, The Ohio State University, Columbus, OH, USA
7 Department of Pediatric Plastic and Reconstructive Surgery, Nationwide Children's Hospital Columbus, OH, USA 8 Division of Dermatology, Nationwide Children's Hospital, Columbus, OH, USA
9 Department of Pathology and Laboratory Medicine, Nationwide Children's Hospital, Columbus, OH, USA
10 Divisions of Genetics and Genomic Medicine, Nationwide Children's Hospital, Columbus, OH, USA

Vascular anomalies and neurocutaneous disorders are caused by inherited and/or post-zygotic variants in the PI3K/AKT/mTOR and RAS/RAF/MEK signaling pathways. Somatic second-hit variants resulting in a complete loss-of-function of a gene product were described by Knudson in the setting of a tumor suppressor. Similarly, Knudson's double-hit mechanism has been proposed to explain the clinical variability in patients with vascular anomalies and neurocutaneous disorders. Second-hit variants are understudied due to the need to sample lesional tissues and tissue heterogeneity. Here we describe the use of high-depth paired exome analysis to discern the genetic underpinnings of these disorders. Comprehensive genomic profiling, including high-depth paired exome sequencing of a comparator germline sample and disease-involved tissue, and whole transcriptome analysis of disease-involved tissue, was performed across a cohort of 61 individuals affected with the described disease entities. Single nucleotide variation (SNV), indels, copy number alterations (CNAs), and copy-neutral loss of heterozygosity (cnLOH) were detected using an established bioinformatic pipeline.

(continue to page 42)

(continued from page 41)

Among this cohort, 40 individuals had clinically significant variants (pathogenic/likely pathogenic or Tier I/II), resulting in a diagnostic yield of 66%. We identified 7 individuals with germline pathogenic variation in which disease-affected tissues harbored second-hit variants, including 5 involving GLMN (glomuvenous malformation), and 1 each involving NF1 (neurofibromatosis), and TSC1 (tuberous sclerosis). Three individuals with GLMN germline variants demonstrated 1p cnLOH in lesional tissues as the second hit demonstrating the breadth of detection of disease-associated variation beyond that of SNVs and indels. We did not detect second-hit variants in the lesional tissues of three individuals with germline pathogenic variation in GLMN, PTEN, and KRIT1. The comparisons of inferred stromal and immune cell fractions derived from expression data between cases with and without the detection of second-hit variants did not show a significant difference. High-depth paired exome sequencing enables the detection of both germline variation and second-hit variants in disease-involved tissue. This group of disorders is phenotypically heterogenous which may be attributed to germline and tissue-specific genotypes and gene expression. Additional studies are needed to further delineate clinical heterogeneity and the relationship to second-hit variants.

ABSTRACT SELECTED TALK

Comprehensive, integrated long-read genomic, epigenomic, and transcriptomic characterization for solving the molecular basis of Mendelian conditions

Andrew Stergachis
University of Washington School of Medicine

Highly accurate long-read sequencing has the potential to dramatically enhance our understanding of the genetic basis of Mendelian conditions through the improved detection of genetic variation (e.g., HiFi sequencing), as well as the elucidation of the functional impact of genetic variation on RNA splicing (e.g., Iso-Seq), and gene regulation (e.g., mCpG detection and Fiber-Seq). To test the utility of these long-read approaches for evaluating individuals with Mendelian conditions, we combined HiFi sequencing, bulk MAS-Seq, and Fiber-seq to enable the simultaneous long-read multi-omic profiling (i.e., genome, CpG methylome, chromatin epigenome, and transcriptome) of patient samples using a single sequencing run.

To accomplish this, we developed a machine learning approach for Fiber-seq data to accurately assemble haplotype-resolved single-molecule chromatin accessibility, nucleosome positioning, and transcription factor occupancy patterns using N6-methyladenine stenciled-chromatin fibers. In addition, we demonstrated that Fiber-seq data retains highly accurate variant calling (SNV F1 99.94%) and the diploid de novo genome assembly capabilities from a single PacBio Revio sequencing run (contig N50s 36.3/28.8 Mb).

We next applied long-read multi-omic profiling to cells from a participant within the Undiagnosed Diseases Network (UDN) with unexplained developmental delay, polymicrogyria, sensorineural hearing loss, and bilateral retinoblastomas, with a known chromosome X-13 balanced translocation of uncertain significance. Long-read genomic data identified that the translocation causes NBEA haploinsufficiency, explaining her developmental delay. Long-read transcriptomic data identified that the translocation also creates a PDK3-MAB21L1 fusion transcript that is predicted to disrupt pyruvate dehydrogenase activity, explaining her features associated with mitochondrial and pyruvate dehydrogenase dysfunction. Finally, long-read epigenomic data identified that the translocated chromosome containing both XIST and RB1 is subjected to X chromosome inactivation (XCI) in ~5% of cells, resulting in the silencing of the RB1 locus that is located 13 Mb away from the translocation breakpoint, thus representing the "first hit" for predisposing her to retinoblastomas. Overall, we demonstrate the utility of long-read multi-omic profiling for resolving the mechanisms underlying unsolved rare conditions.

Copy-number variants as modulators of common disease susceptibility

POSTER 205

Chiara Auwerx[1,2,3,4], Maarja Jõeloo[5,6], Marie C. Sadler[2,3,4], Nicolò Tesio1, Sven Ojavee2,3, Charlie J. Clark1, Reedik Mägi6, Estonian Biobank Research Team[6,§], Alexandre Reymond[1] & Zoltán Kutalik[2,3,4]

1 Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland
2 Department of Computational Biology, University of Lausanne, Lausanne, Switzerland
3 Swiss Institute of Bioinformatics, Lausanne, Switzerland
4 University Center for Primary Care and Public Health, Lausanne, Switzerland
5 Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia
6 Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia
§ Estonian Biobank Research Team: Tõnu Esko, Andres Metspalu, Lili Milani, Reedik Mägi, Mari Nelis

Copy-number variations (CNVs) have been associated with genomic syndromes but their impact on health later in life in the general population remains poorly described. Assessing four modes of CNV action, we performed genome-wide association scans between the copy-number of CNV-proxy probes and 60 manually curated ICD-10 based clinical diagnoses in 331,522 unrelated white UK Biobank participants with replication in the Estonian Biobank. We identified 73 signals involving 40 diseases, all of which increased disease risk and caused earlier onset. Even after correcting for these signals, a higher CNV burden increased risk for 18 disorders, mainly through the number of deleted genes, suggesting a polygenic CNV architecture. Pathogenicity of individual CNV regions was driven by both by the number and nature of encompassed genes, as CNVs encompassing a larger number of genes were more pleiotropic (+0.16 associations per affected gene; p = 1.5 x 10-5) and disease-associated CNVs encompassed more constrained genes (ppLI = 1.3 x 10-4; pLOEUF = 1.9 x 10-7). Associations were supported by colocalization with both common and rare single nucleotide variant association signals, providing insights into the epidemiology of known gene-disease pairs typically studied in clinical cohorts (e.g., BRCA1 and LDLR deletions increase ovarian cancer and ischemic heart disease risk, respectively), clarifying dosage mechanisms of action (e.g., both increased and decreased 17q12 dosage affects renal health), and identifying putative causal genes (e.g., ABCC6 for kidney stones). Characterization of the pleiotropic pathological consequences of recurrent CNVs in adulthood indicated variable expressivity of these regions and the involvement of multiple genes.

(continue to page 45)

(continued from page 44)

For instance, we find evidence for involvement of genes beyond TBX1 for common cardiac phenotypes in 22q11.2 CNV carriers and identify novel non-neurological disease associations with 15q13 CNVs that specifically involve dosage of genes within an interval that does not span CHRNA7, the suggested candidate gene for neuropsychiatric afflictions previously identified in carriers. Together, our results shed light on the prominent role of CNVs in determining common disease susceptibility within the general population and provide actionable insights allowing to anticipate later-onset comorbidities in carriers of recurrent CNVs.

Cost Effective Cancer Detection Using Genome Wide Error Suppressed Sequencing

POSTER 206

Paul Tang

While different features of cell-free DNA (cfDNA) have been utilized for early cancer detection, somatic mutation has been shown to offer the highest specificity for ctDNA detection in plasma. The recent drop of sequencing cost makes it possible to track thousands of somatic variants from the cancer genomes as cancer markers in plasma, achieving very high sensitivity and specificity for molecular residual disease detection. This approach, however, will be limited by errors introduced during library preparation and sequencing. Here, we describe a highly efficient ssDNA sequencing technology, that is able to convert both ssDNA and dsDNA cfDNA molecule to concatemer so that one can effectively identify true somatic variant signal at single molecule level. It is the first NGS technology that combine rolling cycle amplification of circularized cfDNA molecules with concatemer based error correction to achieve high sensitivity and specificity for mutation detection. It can remove polymerase errors from the first round of amplification and retain full molecule information of cfDNA. We deployed this technology in an analytical study with contrived samples and demonstrated a robust sensitivity at 10-5 at a specificity of 98% without the use of any patient specific reagent in enrichment of the genome of interest.

Cross-species systems genetics to identify novel determinants of cardiac function

POSTER 207

Giacome von Alvensleben

Heart failure (HF) is the most prevalent cardiovascular disorder and a leading cause of mortality worldwide. Despite this, the genetic and molecular determinants of HF, particularly those related to diastolic HF which account for half of the cases, are still only partially known. We conceived an integrative multi-omics systems genetics strategy in 32 strains of the BXD mouse genetic reference population, which replicates some of the heterogeneity of the human population. We performed state-of-the art phenotyping on 320 male mice fed with a chow or high fat diet mimicking healthy or unhealthy human dietary habits and collected a large set of clinical systolic, diastolic, and anatomical echocardiography traits. Integration of these data with left ventricle transcriptome, lipidome and proteome profiling allows to explore the impact of the diet, genetic background, and their interaction on the collected cardiac phenotypes. We demonstrated that diet and genotype have a heterogeneous impact on cardiac function in the BXD population, replicating a broad range of cardiac conditions overlapping with human-like heart disease features. Through systems genetics analyses, we uncovered several significant genotype-to-phenotype associations. We mapped hundreds of significant quantitative trait loci (QTL) for transcriptomics, proteomics, lipidomics and phenotypic traits. Of note, this approach unveiled significant QTLs associated with diastolic parameters, including mitral valve (MV) flow A-wave peak velocity, A-wave duration and MV E/A ratio. A gene prioritization pipeline was developed to identify putatively causal candidate genetic determinants of these cardiac phenotypes. To validate our targets, provide clinical relevance and support causality, we are developing a cross-species validation strategy in the UK-Biobank by performing GWAS followed by functional interpretation of the associations. With those results, we will expand the mouse prioritization strategy in humans to identify concordant gene- or pathways-to- phenotype associations across species. We predict that our approach will lead to the identification of novel gene targets influencing cardiac performance with a particular focus on diastolic function and, consequently, the onset of HF. Describing new causal mechanisms or molecular-to-phenotype associations involved in left-ventricle cardiac physiology could contribute to the development of new therapeutic strategies, currently limited for all forms of HF.

Diagnosing Spinal Muscular Atrophy (SMA) from Exome or Targeted Sequencing Data

POSTER 208

Ben Weisburd[1], Rakshya Sharma[2,1], Chrissy Austin-Tse[1], Vijay Ganesh[1,8], Ikeoluwa Osei-Owusu[1], Emily O'Heir[1], Melanie O'Leary[1], Lynn Pais[1], Villem Pata[10,12], Tiia Reimand[10,11], Katrin Õunap[10,11], Sander Pajusalu[10,11], Carsten Bönnemann[13], Sandra Donkervoort[13], Joseph G. Gleeson[3], Göknur Haliloğlu[14], Audrey L. Daugherty[4], Peter B. Kang[4], Gianina Ravenscroft[5], Hamish S. Scott[6], Ana Töpf[7], Anne O'Donnell-Luria [1,9], Grace Tiao[1], Heidi L. Rehm[1]

1 Program in Medical and Population Genetics, Broad Institute of MIT and Harvard / Center for Genomic Medicine, Massachusetts General Hospital, Cambridge, MA, USA
2 UC Santa Cruz Genomics Institute, UCSC, Santa Cruz, CA, USA
3 Department of Neurosciences, University of California, San Diego, La Jolla, CA, USA
4 Department of Neurology, University of Minnesota Medical School, Minneapolis, MN, USA
5 Centre of Medical Research, The University of Western Australia and the Harry Perkins Institute for Medical Research, Perth, Western Australia, Australia
6 Centre for Cancer Biology, An SA Pathology & UniSA Alliance, Adelaide, SA, Australia
7 John Walton Muscular Dystrophy Research Centre, Newcastle University and Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK
8 Department of Neurology, Brigham & Women's Hospital and Harvard Medical School, Boston, MA, USA
9 Department of Pediatrics, Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA, USA
10 Department of Clinical Genetics, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia
11 Genetics and Personalized Medicine Clinic, Tartu University Hospital, Tartu, Estonia
12 Anesthesiology and Intensive Care Clinic, Tartu University Hospital, Tartu, Estonia
13 Neuromuscular and Neurogenetic Disorders of Childhood Section, Neurogenetics Branch, National Institute of Neurological Disorders and Stroke, NIH, Bethesda, MD, USA
14 Department of Pediatrics, Division of Pediatric Neurology, Hacettepe University Faculty of Medicine, Ankara, Turkey

Diagnosing spinal muscular atrophy (SMA) from exome or targeted sequencing data Spinal muscular atrophy (SMA) is a progressive, often fatal genetic disorder of motor neuron degeneration, causing progressive loss of muscle function throughout the body. It affects between 1 in 6,100-11,500 individuals and is the second most common recessive genetic disorder in Europeans. Although highly accurate tools exist for diagnosing SMA from short read genome or long read data, none exist for patients with gene panel or exome sequencing, which is often the first line testing for neuromuscular disorders. We therefore developed SMA Finder - a new tool that can diagnose SMA from exome, genome, and targeted sequencing data with 100% accuracy demonstrated to date. For each sample, SMA Finder evaluates reads that overlap the c.840 positions of the highly homologous SMN1 and SMN2 genes and reports whether the sample has zero functional copies of SMN1 along with a confidence score.

(continue to page 49)

(continued from page 48)

We applied SMA Finder to 21,312 exome and 6,655 genome samples from phenotypically heterogeneous rare disease cohorts within the Broad Institute Center for Mendelian Genomics and the Rare Genomes Project through the GREGoR consortium. SMA Finder showed 100% sensitivity on 13 known SMA-positive samples (10 exomes and 3 genomes) as well as 100% specificity on 11,248 unaffected samples (8,228 exomes and 3,020 genomes). SMA Finder also flagged 11 previously-undiagnosed exome samples as candidate SMA cases, of which five have now been validated by gold standard methods. For the remaining (n=6), further confirmation testing is pending. Strikingly, out of the 24 total SMA-positive cases in our cohorts, 18 had an initial clinical diagnosis of muscular dystrophy and two had a clinical diagnosis of congenital myasthenic syndrome.

To evaluate performance on other cohorts, we also ran SMA Finder on 9,923 samples (1,157 exomes and 8,766 targeted sequencing samples) from a heterogeneous rare disease cohort at the Tartu University Hospital. SMA Finder detected three known cases and three undiagnosed cases which have since been confirmed by gold standard methods. Of the three novel cases, two presented as SMA type III, and one had symptoms that did not include obvious SMA features.

SMA Finder is publicly available at https://github.com/broadinstitute/sma-finder and has proven to be a fast and highly accurate tool for detecting SMA cases in exome, genome, and targeted-sequencing datasets.

Discordances between WHO and ICC classification systems lead to diagnostic and prognostic ambiguity in a subset of 3,499 real-world MDS patients

POSTER 209

Grant Hogg[1], Eric A Severson[1], Michael Biorn[2], Shreegowri Mattighatta Bhojanna[3], Qian Zeng[4], Heidi Hoffmann[4], Li Cai[5], Wenjie Chen[4], Sabrina Gardner[5], Lax Iyer[4], Angela Kenyon[4], Deborah Boles[5], Scott Parker[5], Stanley Letovsky[4], Henry Dong[6], Narasimhan Nagan[4], Shakti Ramkissoon[1,7], Marcia Eisenberg[5], Anjen Chenn[5], Taylor J Jensen[1]

1 Labcorp Oncology, Burlington, NC, USA
2 Labcorp, Marietta, GA, USA
3 Labcorp, Richardson, TX, USA
4 Labcorp, Westborough, MA, USA
5 Labcorp, Durham, NC, USA
6 Labcorp, New York, NY, USA
7 Wake Forest Comprehensive Cancer Center and Department of Pathology, Wake Forest School of Medicine, Winston-Salem, NC, USA

Myelodysplastic syndromes (MDS) are heterogeneous disorders of the blood characterized by dysplasia, blast counts <20% and recurrent genetic abnormalities. In 2022, The World Health Organization (WHO) and International Consensus Classification (ICC) updated guidelines to better define MDS and acute myeloid leukemia (AML) utilizing genetic abnormalities; however, differences exist between the two classification systems. The WHO has retained the 20% blast count cutoff for AML but the ICC has lowered the cutoff to now classify patients with >10% blasts and specific recurrent genetic abnormalities as either AML or the new intermediate entity MDS/AML. We analyzed 3,499 MDS patients for whom Next Generation Sequencing (NGS) panel testing and blast count results were available to assess the potential impact of these changes.

NGS was performed on whole blood or bone marrow samples from 3,499 patients using a panel to detect SNV/Indels across 50 genes. Patients with cause for testing MDS or cytopenias were submitted for NGS analysis as part of routine clinical care. Disease status or symptoms were abstracted from test requisitions for each patient. Bone marrow blast counts were measured up to 90 days before and 14 days after NGS.

(continue to page 51)

(continued from page 50)

139/3499 (4.0%) MDS patients had 10-19% blasts and genetic abnormalities that could result in discordant diagnoses between ICC and WHO. Of these patients, 13/139 (9.4%) had NPM1 alterations and >10% blasts which allows for the ICC diagnosis of "AML with mutated NPM1" but only "MDS with increased blasts" (MDS-IB2) by the WHO. 46/139 (33.1%) patients had TP53 alterations including 15 (10.8%) with single hit TP53 which can diagnose "MDS/AML with mutated TP53" by the ICC but only MDS-IB2 with no further specification by the WHO. 93/139 (66.9%) patients had mutations in genes that help define "MDS/AML with myelodysplasia-related gene mutations" – a subtype also not specified by the WHO.

This study highlights the grey areas between the ICC and WHO classification systems that could affect diagnosis, treatment, clinical trial enrollment, and reimbursement of MDS related care. For prognosis,

patients with MDS/AML with TP53 single mutation could have their prognosis under-estimated by established MDS prognosis schemes such as the IPSS-M that utilize the WHO.

Elucidating the influence of chromatin on plasmid-based reporter assays

POSTER 210

Benjamin J. Mallory[1], Danilo Dubocanin[4], Lea M. Starita[1,2], Andrew B. Stergachis[1,2,3]

1 Department of Genome Sciences, University of Washington
2 Brotman Baty Institute
3 Division of Medical Genetics, Department of Medicine, University of Washington
4 Department of Genetics, Stanford University School of Medicine

Although plasmid-based reporter assays have been the backbone of functional genomics for decades, what exactly happens to plasmids after they are transfected into mammalian cells remains a mystery. Not knowing how chromatin is organized on plasmids relative to the genome and how that influences the activity of regulatory elements (RE) in both contexts limits the biological interpretability of reporter assays. To bridge this knowledge gap, we adapted a single molecule sequencing method called Fiber-seq to detect chromatin architecture along individual plasmid molecules following transfection into mammalian cells.

We find that >90% of transfected plasmid molecules localize to the nucleus where they exist in a wide distribution of chromatin states, ranging from molecules bound by regularly spaced nucleosomes to those completely nucleosome-free. Plasmid-bound nucleosomes appear to result in smaller footprints when compared to their nuclear genome-bound counterparts, suggesting potential differences in nucleosome stability or composition along plasmids. However, over a period of 2-3 days, the plasmid-bound nucleosomes adopt a chromatin state that more resembles that assembled along the nuclear genome. Notably, RE sequences in a plasmid context maintain some canonical characteristics of genomic REs, but this pattern appears at least partially cell-type dependent, highlighting potential cell context dependencies on plasmid chromatinization. Overall, we demonstrate that plasmids are indeed chromatinized upon transfection. However, plasmid chromatinization is highly heterogeneous on a single-molecule level, and diverges from that observed along the nuclear genome.

Evaluation of the T2T-CHM13 as a reference standard for somatic variant Detection

POSTER 211

Zonggao Shi, **David Rosenfeld, Caleb Gallops, Karissa Dieseldorff Jones, Xiaotu Ma, David A. Wheeler**

1 Department of Computational Biology, St Jude Children's Research Hospital, Memphis, TN 38105

The human reference genome sequence is a critical resource for clinical Next-Generation Sequencing (NGS) testing. Although GRCh38 is widely recommended for use in research and clinical NGS tests, more than half of the clinical laboratories still use the GRCh37/hg19 version. The recent introduction of the T2T-CHM13/hs1 reference genome has ignited interest as a potential reference genome due to its near completeness and resultant superior mapping characteristics for short read NGS data. Here we compare hg19 and hs1 short read mapping characteristics and the suitability of hs1 for somatic SNV/Indel detection. Tumor and normal paired whole genome sequencing data (Illumina paired-end, read length 126 bp) from six consented patients were mapped onto hg19 and hs1 reference genomes using BWA. Tools for variant detection included GATK4 (Picard for QC metrics and Mutect2 for SNV and indel calling) and Delly for Copy Number Variation (CNV). BAM alignment statistics were extracted with Alfred. Compared to hg19, mapping with hs1 resulted in slightly less median coverage (89.7 vs 91.2), but with superior alignment metrics such as increased covered bases (3.1 billion vs 2.9 billion), far less variance in coverage deviation (1.1 vs 3.5), a higher fraction of mapped pair reads (0.97 vs 0.94), and a reduced fraction of unmapped reads (0.02 vs 0.05). The improved uniformity in coverage enhanced the resolution for CNV profiling, and reduced artifacts and ambiguity. SNV/Indel calling with Mutect2 on tumor-normal paired samples using hs1 led to 11-22% fewer raw variant counts, mainly through elimination of artifacts due to mapping errors. Although the annotation resources for hs1 are not yet as extensive as for hg19 or GRCh38, our exploratory results demonstrate the potential benefits of an accurate and complete reference genome for NGS read alignment and somatic variant detection.

Fully automated high quality NGS library preparation for all investigators: optimized tagmentation chemistries and workflows

POSTER 212

Shreyas Shah

Advancements in Illumina sequencing methods have necessitated novel approaches for Next Generation Sequencing (NGS) library preparation. In recent years, the library preparation process has undergone biochemical optimizations as well as mechanical approaches for automation. Several library preparation protocols employ tagmentation chemistries, which promise workflow simplification and reduced bias in DNA cleavage sites compared to sonication and enzymatic fragmentation, making it an attractive option for NGS researchers. Our group's previous work has adapted enzymatic fragmentation and sonication-based library preparation assays to a custom platform, integrating in-capillary thermal cycling, fluid transport, and biochemical reactions. In this work we expand on previous strategies by automating a tagmentation-based protocol, addressing several concerns that distinguish tagmentation assay automation from enzymatic fragmentation library preparation assays. As a bead-based chemistry, tagmentation requires specialized liquid handling techniques that ensure optimal reaction conditions. Bead-linked transposomes (BLTs) differ from SPRI purification beads in magnetic behavior, size, surface morphology, and biomechanical reactivity. We have developed custom mixing and liquid handling protocols, as well as a method of manipulating the magnetic field throughout the assay to achieve more homogenous reaction mixtures. The protocol is fully automated at less than six hours of on-BioQule platform run time, with both final library concentration and electropherograms comparable to manual positive controls.

For research use only. Not for use in diagnostic procedures.

Gene-disease validity recuration efforts of the ClinGen Hearing Loss Expert Panel

POSTER 300

Kezang C. Tshering1, Marina T. DiStefano1,2, Andrea M. Oza1,2, Pamela Robertson1, Ryan Webb1, Enyonam Edoh1, Ellie Broeren1, Julie Ratliff1, Vanessa Gitau1, Heidi L. Rehm, Ahmad N. Abou Tayoun3, and Sami S. Amr2 on behalf of the ClinGen Hearing Loss Clinical Domain Working Group

1 The Broad Institute of MIT and Harvard, Cambridge, MA, USA
2 Laboratory for Molecular Medicine, Partners Healthcare Personalized Medicine, Cambridge, MA, USA
3 Al Jalila Children's Specialty Hospital, AL Jaddaf, Dubai, United Arab Emirates

The Clinical Genome Resource (ClinGen) supports an international network of experts in building an authoritative resource that defines the clinical relevance of genes and variants for use in precision medicine and research. For genetic testing to lead to proper disease diagnosis, gene-disease relationships must be transparently and systematically evaluated. The ClinGen Hearing Loss Gene Curation Expert Panel (HL GCEP) was one of the first GCEPs to be assembled in 2016, and has since curated 174 gene-disease relationships using ClinGen's semi-quantitative framework. However, evidence is constantly accumulated for all genetic diseases as new patients are diagnosed, and general refinements to the ClinGen curation guidelines may change the classifications over time. In order to keep the curations up-to-date, ClinGen requires all genes with Disputed, Limited, Moderate, and Strong relationships to be recurated every 2–3 years, whereas, gene-disease relationships classified as Refuted and Definitive are curated only if a counter argument arises. The HL GCEP recuration process involved re-evaluating all previous genetic and experimental evidence of numerous genes using the current SOP and conducting a full literature search for any new evidence on their gene-disease relationships. The recurations were presented to the GCEP for re-approval and once approved, the recurated evidence and the evidence summaries were updated and published to the ClinGen website (www.clinicalgenome.org). From 2022 through summer of 2023, the GCEP recurated 36 gene-disease pairs, of which 8/36 (22%) changed their classification. Unchanged were 2 Disputed, 17 Limited, 8 Moderate and 1 Strong. Two Moderate and 5 Strong genes upgraded to Definitive and 1 Strong gene downgraded to Limited. Across the 8 genes that changed their classification, new evidence led to an upgrade and the downgrade was due to re-evaluation of the existing evidence using updated ClinGen gene curation guidance. Across all ClinGen GCEPs, gene recuration is an ongoing effort and expert-reviewed classifications are posted on an ongoing basis. The high rate of change of the hearing loss genes in this study emphasizes the importance of ongoing recuration efforts to ensure genetic tests and research efforts are applying the most up-to-date evidence in their diagnoses and studies.

Genomic approaches for disease mechanism & biomarker discovery in Dry Eye Disease

POSTER 301

Lena Wyss[1], Federica Storti[1], Fabian Koechl[2], Vera Griesser[2], Roland Schmucki[3], Martin Ebeling[3], Nils Schärer[4], Zisis Gkatzioufas[4], Christoph Ulmer[1], Philip Knuckles [2]

Roche Pharma Research and Early Development
1 I2O - Ophthalmology
2 PS - BioX Genetics & Genomics
3 PS- Predictive Modeling and Analytics
4 AugenKlinik, Universitätsspital Basel

Dry eye disease (DED) is a highly prevalent heterogeneous condition. It is a multifactorial disease of the ocular surface characterized by a persistently deficient and/or unstable tear film causing discomfort, inflammation and, in extreme cases, visual impairment. A current challenge in DED care is an incomplete understanding of disease mechanisms as well as specific biomarkers to effectively stratify and treat patients. A minimally invasive method to biopsy the eye surface is Conjunctival Impression Cytology (CIC) which allows the isolation of nucleic acid and live cells from the eye surface.

We have implemented advanced next generation sequencing (NGS) methods to profile CIC samples from DED patients and healthy donors. We have optimized conditions for biopsy preservation and banking that ensure high quality RNA quality and bulk sequencing results. We also established a robust single-cell RNAseq (scRNAseq) workflow for CIC biopsies and benchmarked cell populations recovered against flow cytometry results. We applied a multiplexing strategy using lipid-conjugated oligo tagging of individual samples and pooling. This approach dramatically increases sample throughput (up to12X) and dramatically reduces labor and reagent costs.

Comparison of a limited cohort of DED samples to healthy individuals reveals highly relevant gene-expression changes in affected conjunctiva such as an upregulation in T-cell specific markers. scRNAseq datasets show high quality cells with expected identities (eg. Epithelial and T-cell subsets). The proportions of cell populations are highly concordant to those identified by FACS analysis. Interestingly, we identified rare and highly disease-relevant cell-types such MITF+ Melanocytes and HEPCAM2+ Goblet cells, not detectable with orthogonal cell detection methods.

Taken together we have established a powerful genomics-based tool set to molecularly profile the surface of the eye. Within the context of DED, this methodology will pave the way in improving our understanding of this highly heterogeneous disease and has wide applicability to other anterior eye pathologies.

Genome British Columbia: beyond 23 years of investing in genomics research and innovation

POSTER 302

Chen Wan[1], Ph.D.

1 Genome British Columbia, Vancouver, British Columbia, Canada

Since its inception in 2000, Genome British Columbia (Genome BC) has been an independent non-profit organization dedicated to applying the power of genomics to pressing societal, environmental, and economic challenges. We support world class genomics research and innovation with the aim of growing a globally competitive life science sector and delivering sustainable benefits for BC, Canada and beyond. We have leveraged our strategic programs and projects to manage a cumulative portfolio of nearly $1.3B in over 500 research projects, technology platforms and innovation initiatives. This includes 1,117 collaborations with partners across BC, Canada and in 42 countries, $969M in cofounding secured, 152 BC-based companies advanced, 3,718 scientific papers published and 767 patent applications filed.

We have a mandate to apply the power of genomics to better the lives of British Columbians and all Canadians through a high performing health care system and to realize precision health. Genome BC will drive the responsible uptake of genomics by bridging the gap from research to the clinic. Through better disease prevention, diagnosis and treatment, genomics will improve health outcomes and sustainability of BC's health care system. Through genomics we will develop research, data, capacity and technology to deploy new tools and knowledge and we will do this with a deliberate focus on access, equity and education.

We strategically promote data collection, management, storage, analysis, and integration – all of which are essential to the optimization of genomic research and innovation. Our data strategy represents a cross cutting pillar in our research and innovation mandate and builds on the notion that publicly funded data are public goods and as such should be openly shared to deliver social, environmental, and economic benefits. To effectively implement genomics technology in the clinic, experts not only need to be aware of the tools available and have clear access to them, but they must also understand the utility. Genomic Education for Health Professionals program is the initiative supporting the development of new genomic educational tools for BC health professionals with a goal of increased equitable access to genetic and genomics services for all British Columbians.`

Highly accurate assembly polishing with the encoder-only transformer DeepPolisher

POSTER 303

Mira Mastoras[1], Mobin Asri[1], Benedict Paten[1], Kishwar Shafin[2]

1 UC Santa Cruz Genomics Institute, Santa Cruz, CA, USA
2 Google LLC, 1600 Amphitheatre Pkwy, Mountain View, CA, USA

Faithfully reconstructing the genome of a species is essential to studying its underlying biology and mechanisms which lead to disease. However, even the highest quality diploid assemblies retain base-level errors, primarily in small indels in low complexity repeats and heterozygous variants in long stretches of homozygosity. Improving the accuracy of these regions is critical to prevent researchers from conflating true biological function with technical artifacts. Current polishing tools are not equipped to fix these difficult remaining errors, however. They rely on statistical heuristics that cannot model the complexity of the error process, which is influenced by genomic sequence context, biases in the read alignment algorithms, read errors from the sequencer and any preceding processing tools. Variant callers utilizing convolutional neural networks have proven they are able to learn this complexity, yet they are not the correct model for polishing because they are trained to classify a genotype, not correct a sequence. Here we present an encoder-only transformer model, DeepPolisher, which predicts the entire sequence in a 100bp window from Pacbio HiFi read alignments. DeepPolisher is the appropriate model for the task of assembly polishing because it is trained to predict a sequence while taking into account the complexity behind read alignment ambiguities. Our pipeline also includes a method, PHARAOH (Phasing Reads in Areas Of Homozygosity), which ensures alignments are accurately phased and allows us to correctly introduce heterozygous edits to falsely homozygous regions. We demonstrate that the DeepPolisher pipeline can take the HPRC HG002 y1 assembly from 7.7 errors per megabase of sequence to 3.9 errors per megabase, with a greater than 60% reduction in indel errors. For the same assembly in the Genome in a Bottle high confidence regions we demonstrate a QV (Quality Value) improvement of 67 to 70, when calculated with both Illumina and HiFi kmers. Our next steps in adapting DeepPolisher to different sequencing technologies, such as ONT, promises to translate these improvements to lower coverage assemblies in projects aiming to leverage the power of long-read sequencing for large scale gene association studies.

Image-based profiling to identify chemical modulators of metabolic disease-relevant cell states

POSTER 304

Felipe R. C. dos Santos[1,2], Hesam Dashti[1,2,3,4], Patrick Byrne[1], Masha Alimova[1], Adam Stefek[1], Lizzie McGonagle[1], Phil Kubitz[1,2], Yu Han[1], Niranj Chandrasekaran[1], Erin Weisbart[1], Beth Cimini[1], Shantanu Singh[1], Anne Carpenter[1], Melina Claussnitzer[1,2,3,4]

1 Broad Institute of MIT and Harvard, Boston, MA
2 The Novo Nordisk Foundation Center for Genomic Mechanisms of Disease, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
3 Department of Medicine, Harvard Medical School, Boston, MA
4 Center for Genomic Medicine and Endocrine Division, Massachusetts General Hospital, Boston, MA

The individual risk of cardiometabolic disease is modified by changes to the mass, distribution and function of adipose tissue and large-scale genetic association studies have robustly linked genetic variants to target genes implicated in adipocyte morphology and function. Image-based cellular profiling is a powerful tool to systematically discover regulators of biological processes in adipocytes in an unbiased manner, and test how metabolic disease-relevant stimulatory conditions and compounds modulate these processes. We first studied undifferentiated pre-adipocytes and the response effect of six disease-relevant stimulatory conditions (isoproterenol, low glucose, rosiglitazone-free, hypoxia, and a combination of interleukin-6 and TNF-alpha) in differentiated adipocytes using image-based profiling. The data revealed that conditions cluster by adipocyte cell differentiation stage and disease stimulatory condition. After stringent quality control and data normalization we performed PCA and clustering analyses, using ~4,400 features and observed that image-based profiles cluster into meaningful pathways. Using the cell-based profiles from these stimulations, our accomplishments are threefold: (1) we established metabolic disease-relevant cell states using image-based profiling in response to stimulatory conditions; (2) we identified compounds that rescue or exacerbate these metabolic disease states; (3) we validated the effect for some of the compounds by using aggregated genome-wide polygenic risk scores (PRSs). Specifically, after mapping the stimulation-responsive disease states in adipocytes, we studied the effect of 300 compounds from the Broad's Target2 drug repository compound library in combination with these five stimulatory conditions.

(continue to page 60)

(continued from page 59)

Moreover, we found compounds with known and novel effects on metabolic disease states. For instance, the data revealed that the compound AZ191 – targeting the gene DYRK1B, which harbors rare gain-of-function mutations associated with abdominal obesity, insulin resistance and type 2 diabetes – rescues the metabolic disease phenotypic state in adipocytes. Finally, we mapped compound effects on the metabolic-relevant states to lipodystrophy PRS associated image-based profiles. We retrieved the direction of the connectivity between them, and identified sets of compounds that rescue and exacerbate lipodystrophy polygenic signatures, respectively. These results provide a solid framework for mapping disease stimulatory cell states, compound effects, cell line reproducibility, and polygenic effects using image-based profiling to learn deep mechanistics insights for metabolic disease.

Immune profiling tools in whole genome and transcriptome analysis of metastatic tumours inform immunotherapy clinical trial design

POSTER 305

Kathleen Wee[1] , Erin Pleasance[1] , Emma Titmuss[2] , Laura Williamson[1] , Yaoqing Shen[1], Karen Mungall[1], Eric Chuah[1], Andrew Mungall[1] , Melika Bonakdar[1], Greg Taylor[1], Veronika Csizmok[1], Cameron J Grisdale[1] , Richard D Corbett[1], Jessica Nelson[1], Stephen Yip[3], Sophie Sun[2], Howard Lim[2], Daniel Renouf [2], Steven JM Jones[1,4], Marco A Marra[4], Janessa Laskin[2]

1 Canada's Michael Smith Genome Sciences Centre, BC Cancer, Vancouver, BC, Canada
2 Department of Medical Oncology, BC Cancer, Vancouver, BC, Canada
3 Department of Pathology and Laboratory Medicine, Vancouver General Hospital, Vancouver, BC, Canada
4 Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada

Advances in sequencing technologies and our understanding of cancer biology and driver events are improving our ability to deliver precision medicine to cancer patients. The rapid increase in the use of immunotherapy in cancer treatment has led to a great need for understanding markers of response and resistance. Incorporating immune profiling tools into whole genome and transcriptome sequencing and analysis (WGTA) enables a better understanding of the immunological landscape in metastatic cancer, which can be used to inform clinical trial design. Over 1,000 patients with advanced/metastatic cancers of diverse tumour types were enrolled in the Personalized OncoGenomics (POG) program and underwent WGTA, with findings discussed at a Molecular Tumour Board and used to guide therapy choice. This includes DNA-based data such as mutations, copy number alterations, and mutation signatures, and RNA-based data, including gene expression and fusions. Embedded within the WGTA pipelines are tools which enable deconvolution of the immune microenvironment (CIBERSORT), identification of patient HLA alleles (OptiType), T- and B-cell repertoire (MiXCR) and expression of neoantigens (NetMHCpan). Developing software containers for immune deconvolution (RICO) and measurement of tumour mutation burden (TMBur) can make pipelines consistent and portable, which is imperative when running clinical trials across multiple centres. Incorporating these immunophenotyping tools in our precision oncology pipeline and investigating the response to immune checkpoint inhibitor treatment led to the discovery of known and emerging biomarkers that predicted response more effectively as a combination and resulted in the creation of the Canadian Atezolizumab Precision Targeting for Immunotherapy Intervention (CAPTIV-8) trial.

(continue to page 62)

(continued from page 61)

This phase 2 clinical trial uses multiple WGTA findings, which we named the immune burden variant (IBV) score, to select patients predicted to be most likely to respond to atezolizumab irrespective of tumour type (NCT04273061). To date, ~22% of POG patients have IBV scores that meet the eligibility criteria for the CAPTIV-8 trial, including a variety of tumour types, and it remains open to recruitment. This demonstrates the utility of integrating diverse tools and WGTA analysis to aid in clinical trial design and investigate key factors which are likely to impact response to cancer treatment.

Immune-Related Long Non-Coding RNA Signatures for Tongue Squamous Cell Carcinoma

POSTER 306

Daniel Hu and Diana Messadi

School of Dentistry, University of California, Los Angeles, CA 90095

Tongue squamous cell carcinoma (TSCC) represents one of the major subsets of head and neck cancer, which is characterized by unfavorable prognosis, frequent lymph node metastasis, and high mortality rate. The molecular events regulating tongue tumorigenesis remain elusive. In this study, we aimed to identify and evaluate immune-related long non-coding RNAs (lncRNAs) as prognostic biomarkers in TSCC. The lncRNA expression data for TSCC were obtained from The Cancer Genome Atlas (TCGA) and the immune-related genes were downloaded from the Immunology Database and Analysis Portal (ImmPort). Pearson correlation analysis was performed to identify immune-related lncRNAs. The TCGA TSCC patient cohort was randomly divided into training and testing cohorts. In the training cohort, univariate and multivariate Cox regression analyses were used to determining key immune-related lncRNAs, which were then validated through Cox regression analysis, principal component analysis (PCA), and receiver operating characteristic (ROC) analysis in the testing cohort.

Six immune-related signature lncRNAs (MIR4713HG, AC104088.1, LINC00534, NAALADL2-AS2, AC083967.1, FNDC1-IT1) were found to have prognostic value in TSCC. Multivariate and univariate cox regression analyses showed that the risk score based on our six-lncRNA model, when compared to other clinicopathological factors (age, gender, stage, N, T), was an important indicator of survival rate. In addition, Kaplan–Meier survival analysis demonstrated significantly higher overall survival in the low-risk patient group than the high-risk patient group within both training and testing cohorts. The ROC analysis indicated that the AUCs for 5-year overall survival were 0.790, 0.691, and 0.721, respectively, for training, testing, and entire cohorts. Finally, PCA analysis demonstrated that the high-risk and low-risk patient groups presented significant deviation regarding their immune status.

In summary, a prognostic model based on six immune-related signature lncRNAs was established. This six-lncRNA prognostic model has clinical significance and may be helpful in the development of personalized immunotherapy strategies.

iPSC-SGE: assessing variant effects in differentiated cell types at scale

POSTER 307

Shawn Fayer1, Clayton E. Friedman2,3,4, Sriram Pendyala1, Moez Dawood5, Daniel Yang2,3,4,6, Douglas M. Fowler1,7, Lea M. Starita1,5

1 Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA
2 Institute for Stem Cell and Regenerative Medicine, University of Washington, School of Medicine, Seattle, WA 98109, USA
3 Center for Cardiovascular Biology, University of Washington, Seattle, WA 98109, USA
4 Department of Medicine/Cardiology, University of Washington, Seattle, WA 98109, USA
5 Brotman Baty Institute for Precision Medicine, Seattle, WA 98195, USA 6Cardiology/ Hospital Specialty Medicine, VA Puget Sound HCS, Seattle, WA 98108, USA
7 Department of Bioengineering, University of Washington, Seattle, WA 98195, USA

Multiplexed assays of variant effect (MAVEs), where the functional effects of thousands of variants of a gene are measured simultaneously, are poised to have a significant impact on precision medicine. We have previously shown that up to 70% of variants of uncertain significance identified by genetic testing can be resolved with the addition of functional data derived from MAVEs. However, these assays have not been performed on the many clinically relevant genes that are exclusively expressed in differentiated cell types and not in the utilitarian cancer cell lines that are currently used for MAVEs. To assess variants in these genes, we developed iPSC-SGE, a new saturation genome editing (SGE) method that introduces single nucleotide variants (SNVs) into a single allele of diploid human induced pluripotent stem cells (iPSC). We applied iPSC-SGE to the cardiac myosin binding protein C3 (MYBPC3) gene, in which pathogenic variants are the most common cause of familial hypertrophic cardiomyopathy. We focused on the C10 domain, spanning exons 32-34, which anchors MYBPC3 within the sarcomere and is a hotspot for pathogenic missense variants. We introduced 498 SNVs into exon 32 and the flanking intronic regions of MYBPC3 in iPSCs and differentiated these cells to cardiomyocytes for phenotyping. We used a FACS assay to score variants based on their abundance relative to the wild type allele, a proxy for the ability of MYBPC3 to incorporate into the sarcomere. We scored 492 of the 498 possible variants for function and 79 variants showed reduced function including all 11 pathogenic control variants. We are currently applying iPSC-SGE to the remainder of the MYBPC3 C10 domain and will use our results to inform clinical variant classification. iPSC-SGE massively expands the phenotypic space accessible to MAVEs. For example, the method could be applied to many of the 22 actionable cardiomyopathy risk genes, as well as many other clinically important genes where assessment of variant phenotype in the correct cell context is required.

JBrowse 2 and Apollo 3: Tools for browsing and annotating personalized genomes

POSTER 308

Scott Cain[1], Colin Diesh[2], Robert Buels[2], Garrett Stevens[2], Caroline Bridge[1], Lincoln Stein[1], Ian Holmes[2]

1 Adaptive Oncology, Ontario Institute for Cancer Research, Toronto, Canada
2 Department of Bioengineering, University of California, Berkeley, USA

The examination of structural variations plays a crucial role in comprehending individual human genomes, particularly in the context of disrupted tumor genomes. To address limitations in visualizing complex structural variants, the JBrowse 2 genome browser platform incorporates specialized features to facilitate the visualization of structural variants alongside other data types. JBrowse 2's plugin infrastructure also enables third party developers to write plugins that serve various components of cancer research, such as with the Isoform Inspector plugin.

We are also developing a platform that allows annotating genomic variants with Apollo 3, a plugin for JBrowse 2. Apollo 3 enables researchers to edit gene annotation structure and function, and we are creating features that would be relevant for annotating personalized genomes and variants. JBrowse 2 is uniquely positioned for browsing personalized genomes with comparative views, which allows viewing structural variation such as deletions, inversions, insertions, and translocations in a derived genome with synteny-style views, and we envision that browsing the personalized genome will allow specific annotation of deleterious variants in the context of disease.

Both Apollo 3 and JBrowse 2 are freely available to the public as open source software (https://jbrowse.org). They are developed entirely in JavaScript, and can be run either as a desktop application or as a static website. The JBrowse website provides comprehensive documentation with sections tailored for users, administrators (who can set up JBrowse on their web servers), and developers (who can enhance JBrowse's functionalities using plugins).

Leveraging well-calibrated functional data and computational predictions to resolve variants of uncertain significance

POSTER 309

Malvika Tejura[1], Shawn Fayer[1], Abbye McEwen[1,2,4], Douglas M. Fowler[1,2,3], Lea M. Starita[1,2]

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

Variants of uncertain significance (VUS) are a key barrier to the implementation of genome-guided precision medicine. Currently, more than 1 million, or nearly 90% of the missense variants discovered so far by genetic testing are classified as VUS because there is insufficient evidence to classify them as pathogenic or benign. Functional data and computational variant effect predictors can be used as evidence for variant classification and can scale to the size of the VUS problem, but have had a limited impact because it is difficult to determine how much weight the data could or should carry as evidence.

Recently, ClinGen devised a framework to calibrate functional data for individual genes (Brnich et al., Genome Medicine, 2019) and computational variant effect predictors genome-wide (Pejaver et al., AJHG,2022). We assessed the genome-wide calibration of computational variant effect predictors and found that it obscures gene-by-gene variability in predictor performance and leads to incorrect variant classifications. We propose a novel approach to calibrating computational data that accounts for the variability of computational predictions across genes. We show that for some genes, the ClinGen genome-wide calibrations are accurate. For other genes, gene-specific calibration provides stronger evidence for variant classification than genome-wide calibration. Most importantly, predictions for some genes are inaccurate and therefore should not be used.

Following Clingen's guidance for calibrating and using functional data, we and others have shown that incorporating functional evidence from multiplexed assays of variant effect (MAVEs) can drive the reclassification of many VUS (69% VUS in TP53, 49% for BRCA1, 15% for PTEN, 90% for DDX3X, and 75% for MSH2). The combination of calibrated functional data and correctly calibrated computational predictions have the potential to resolve the millions of VUS plaguing clinical genetics. For some variants, these two pieces of evidence alone would be sufficient, yielding a classification before the variant is encountered in the clinic. Thus, we are calibrating and combining these data types to address the VUS problem.

Long-read characterization of cancer genome rearrangements using Severus

POSTER 310

Ayse Keskus[1],Tanveer Ahmad[1],Ataberk Donmez[1],Asher Bryant[1], Isabel Rodriguez[2], NicoleM.Rossi[2], Yi Xie[2], Byunggil Yoo[5], Rose Milano[2], Hong Lou[3], Jimin Park[4], Joshua Gardner[4], Brandy McNulty[4], Karen Miga[4],Midhat Farooqi[5],Benedict Paten[4],Michael Dean[2], Mikhail Kolmogorov[1]

1 Center for Cancer Research, NCI, Bethesda, MD, USA;
2 Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD, USA;
3 Leidos Biomedical Research, Inc., National Laboratory for Cancer Research, Frederick, MD, USA
4 UC Santa Cruz Genomics Institute, Santa Cruz, CA, USA
5 Children's Mercy Hospital, Kansas City, MO, USA

Structural variation (SV) is one of the hallmarks of cancer, ranging from simpler deletions and amplifications to catastrophic events involving shuffling megabase-scale fragments from one or multiple chromosomes. Recent pan-cancer studies showed that more than 50% driver mutations are explained by SVs. Compared to SNPs and small indels, SVs are much more difficult to detect with the traditional sequencing technologies. As a result, most current cancer genomic studies focus on small variations inside protein-coding regions.

Long-read sequencing can overcome the limitations of the short-read technologies and was recently used to produce the most complete and accurate de novo genome assemblies to date. However, most current long-read assembly and variation calling algorithms were not designed for analysis of tumor genomes and fail to capture somatic variants, complex rearrangements and clonality.

Here we present Severus to detect and annotate a wide range of germline and somatic SVs from long-read tumor sequencing, including complex rearrangements such as chromothripsis and breakage-fusion-bridge. Severus can work in single-sample, tumor/normal, or multi-site modes. The algorithm takes advantage of long-read phasing to produce haplotype-specific calls. Severus also utilizes a genome graph approach to reveal clusters of breakpoints and reconstruct the derived structure of tumor chromosomes.

(continue to page 68)

(continued from page 67)

Severus outperformed the current state-of-the-art tools for germline SV calling on the manually curated set of SVs for the COLO829 tumor cell line. Further, we sequenced five tumor/normal cell lines with ONT and confirmed the high reliability of Severus calls, including chromothripsis, chromoplexy, breakage-fusion-bridge (BFB), inversions with amplifications/deletions, and intrachromosomal long insertions. We also analyzed several HPV-infected cervical cancer cell lines, revealing mosaic amplifications and BFBs near the HPV integration sites. Interestingly, BFBs were exclusive to chromothripsis in almost all cases.

To evaluate the diagnostics potential of the long-read method, we applied Severus to several clinical samples of pediatric leukemia sequenced with long reads. As expected, we only found a few somatic SVs per sample, compared to hundreds in adult cancers. Some SVs were clinically relevant, for example a mosaic chromosome translocation resulting in a MLLT10-KMT2A fusion, which was not found by any of the standard assays.

Machine learning approach using RNAseq for diagnostic classification of pediatric hematologic and non-CNS solid tumors

POSTER 311

Alexander M. Gout1, Daniel Putnam1, Delaram Rahbarinia1, Meiling Jin1, Xiaotu Ma1, Jinghui Zhang1, David A. Wheeler1, Larissa V. Furtado2, Xiang Chen1

1 St. Jude Children's Research Hospital, Department of Computational Biology, Memphis, Tennessee, USA
2 St. Jude Children's Research Hospital, Department of Pathology, Memphis, Tennessee, USA

Molecular features of tumors are increasingly used in the diagnostic classification of pediatric cancer and have become a key component of personalized therapy. Although machine learning classifiers using methylation data have achieved notable success in pediatric central nervous system (CNS) tumor diagnostics, bulk whole transcriptome sequencing (RNAseq) has been less commonly utilized as a diagnostic approach. Here we report the development of a supervised machine learning pipeline to classify pediatric tumors using whole transcriptome expression data. Samples were derived from both research cohorts and routine clinical testing performed at St. Jude Children's Research Hospital on patients consented for research. We harmonized diagnosis subtype annotations for 2395 hematologic malignancy and 1195 non-CNS solid tumor samples guided mainly by WHO criteria, and input from St Jude pathologists. A total of 37 hematologic cancer subtypes (B-ALL (n=16), T-ALL (n=8), and AML (n=13)) and 13 distinct solid tumor subtypes were included for classifier development. We divided the samples into training and test cohorts. Multiple feature reduction algorithms from raw gene-level count data were evaluated. An ensemble classifier combined input from five machine learning algorithms followed by accuracy assessment via bootstrapping on the training set. A confidence threshold was selected to achieve prediction of 95% of the bootstrap samples. We achieved a median accuracy of 96% and 100% on the hematologic and solid tumor test samples respectively. Additional validation on NCI-TARGET RNAseq, an external pediatric multi-subtype cohort, achieved overall median accuracies of 99% (hematologic) and 99% (solid tumor). We could reclassify 80% of cases initially categorized as NOS (not otherwise specified) for hematologic malignancies, and 94% of rhabdomyosarcoma (NOS) into one of our diagnostic categories. These results suggest that machine learning applied to bulk RNAseq may represent a viable approach to aid in tumor diagnosis and clinically meaningful stratification of tumor types.

Mapping the molecular events of early embryogenesis using high-throughput omics and gastruloid models

POSTER 312

Riddhiman Garge

Congenital malformations are a leading cause of infant mortality, however, research efforts to study birth defects have been comparatively limited. Aside from the obvious ethical and experimental challenges associated with obtaining and studying human embryos, models for understanding the dynamic gene regulation underlying human embryonic development have only recently been developed. In vitro, stem cell models called gastruloids have emerged as valuable surrogates for studying early embryonic development. Gastruloids are 3D aggregates of embryonic stem cells that recapitulate the organization of implanted embryos and consist of neural, gut, somite, and brain cell types.

We have taken a multi-omics approach to comprehensively map molecular changes throughout gastruloid development. We combine RNA sequencing and quantitative mass spectrometry to investigate temporal gene expression changes at the level of RNA, protein, and post-translational modifications. We quantify ~8000 proteins across major gastrulation stages including pre- and post-implantation to reveal temporal dynamics of protein expression, the extent of coregulation of the proteome and transcriptome, and the dynamic phosphorylation states of developmentally regulated proteins. Further, by expanding the scope of these multi-omic assays to mouse gastruloids, we systematically compare the divergence in gastrulation programs across mammalian development. Our multi 'omic atlas provides a foundational resource for understanding early human embryonic development at the molecular level and offers a starting point to understand genes associated with congenital defects. Future work will measure the impact of perturbing key developmental genes that underlie Mendelian disease on gastruloid cell type composition and on gene expression programs, creating a phenotyping platform that will provide a mechanistic understanding of disease states in embryonic development.

MedGenome's genomics solutions for precision medicine

POSTER 400

Marco Corbo[1], **K. Suryamohan**[1], **R. Gupta**[2], **R. Menon**[3], **M. Muraleedharan**[1], **E. Ramesh**[1], **H. Gowda**[1], **H. G.H**[1], **V. Ramprasad**[4], **S. Seshagiri**[1]

1 MedGenome Inc., Foster City, CA
2 MedGenome Labs Ltd., Bangalore, India
3 MedGenome Labs Pvt ltd, Bangalore, India
4 MedGenome Inc., Bengaluru, India

Genomics has revolutionized biomedical research, transforming our understanding of human health and disease. MedGenome, a leading provider of sequencing and bioinformatics solutions, offers a comprehensive range of services, including whole genome and whole exome sequencing, transcriptome sequencing, liquid biopsy, immune repertoire analysis, single-cell sequencing, and organoid technology. Our proprietary solutions power pre-clinical research, biomarker discovery, and early drug development pipelines for academia, pharmaceutical and biotechnology companies. With state-of-the-art laboratories in India and the US, our team of highly skilled and experienced scientists can process a wide array of sample types using established best practices to generate high-quality sequencing data. MedGenome has developed an array of cutting-edge bioinformatics tools and optimized wet lab workflows for research and diagnostics. Our cloud-based analytics solution empowers scientists in precision medicine research by facilitating the analysis of multi-omic data to understand disease mechanisms and develop targeted therapies. Our platform can seamlessly analyze large genomic datasets and can generate publication-ready figures and reports.

MedGenome has spearheaded the development of several proprietary technologies including VarMiner, an AI-enabled variant interpretation software, and HiTmAb, an innovative high throughput antigen-specific monoclonal antibody discovery platform that harnesses the power of single-cell B-cell receptor (BCR) sequencing for identification of antigen-specific B-cell antibodies.

MedGenome has also positioned itself as a frontrunner in human genetics research. MedGenome serves over 4000 hospitals and offers more than 1300 genetics tests in India. We routinely generate large-scale genomic data to identify novel human genetic variants to enable drug discovery and diagnostics. As a founding member of the GenomeAsia 100K project, we have curated an extensive catalog of South Asian population-level genomic sequencing data.

(continue to page 72)

(continued from page 71)

This is a novel resource for precision medicine research that enables identification of genetic risk factors and rare variants among underrepresented populations. Taken together, our expertise in genomics, diagnostics, sequencing, and bioinformatics positions us at the forefront of biomedical research, driving groundbreaking advancements and empowers scientists to address complex genomic questions.

ABSTRACT SELECTED TALK

Multi-omics for precision medicine in preeclampsia in an admixed population

Kathryn J. Gray , Juan P. Casas , Norma C. Serrano Díaz , Claudia C. Colmenares ,Babatunde Akinwunmi , Jie Hu , Vesela Kovacheva , David E. Cantonwine , S. Ananth Karumanchi , Richard Burwick , G. Larry Maxwell , Thomas F. McElrath , Richa Saxena

1 University of Washington, Department of Obstetrics & Gynecology, Seattle, WA, USA
2 Brigham and Women's Hospital, Division of Aging, Boston, MA, USA
3 Fundación Cardiovascular de Colombia, Santander, Colombia
4 Brigham and Women's' Hospital, Department of Obstetrics & Gynecology, Boston, MA, USA
5 Massachusetts General Hospital, Center for Genomic Medicine, Boston, MA, USA
6 Cedars-Sinai Medical Center, Department of Medicine, Los Angeles, CA, USA
7 San Gabriel Valley Perinatal Medical Group, Monterey Park, CA, USA
8 Inova Health System, Falls Church, VA, USA
9 Broad Institute, Program in Medical and Population Genetics, Cambridge, MA, USA

Preeclampsia (PE) is a severe pregnancy-specific disorder mediated by the placenta that affects 5% of all pregnancies and is characterized by the development of new-onset hypertension and proteinuria after 20 weeks gestation. PE remains a leading cause of maternal and neonatal morbidity and mortality with significant disparities in prevalence and outcomes between populations. PE has substantial heritability, estimated at 55-60%, with both maternal and fetal contributions, 30-35% and 20%, respectively. However, the underlying etiology of PE remains poorly understood; consequently, predictive and therapeutic options remain limited and delivery is the only cure. To address these challenges, we created the TOPMed Boston-Colombia Collaborative Adverse Pregnancy Outcome study (BCC-PREG) built on several multi-ethnic pregnancy cohorts including LIFECODES (Boston) and GenPE (Colombia) with high-quality biologic samples and extensive demographic and clinical phenotyping. Similar to prior results in European populations, polygenic risk score analysis using GenPE maternal PE GWAS summary statistics demonstrated that maternal genetic predisposition to hypertension, obesity, and type 2 diabetes was associated with PE risk (systolic blood pressure OR 1.28 (95% CI 1.18-1.38), diastolic blood pressure OR 1.22 (1.14-1.30)), BMI (OR 1.15 (1.03-1.27), and type 2 diabetes (OR 1.12 (1.02-1.23)), and inversely correlated with offspring birthweight (OR 0.91 (0.86-0.97)). In multi-ancestry GWAS meta-analysis of GenPE with available maternal PE GWAS data (N=17,007 cases; 37,032 controls), 4 genome-wide significant loci ($p<5x10-8$) and 5 additional suggestive loci ($p<1x10-6$) were identified. Now as part of TOPMed, more than 14,500 maternal and fetal samples from BCC-PREG have undergone whole genome sequencing (WGS) with additional multi-omic profiling (metabolic, proteomic, transcriptomic) maternal plasma samples forthcoming.

(continue to page 74)

(continued from page 73)

Together, BCC-PREG will be the largest pregnancy cohort with WGS to date, with the potential for delineation of novel common and rare maternal and fetal genetic variants, as well as maternal-fetal genomic interactions, underlying the causal biologic pathways leading to PE and its subtypes. The inclusion of both US and Colombian cohorts will additionally inform understanding of disease etiology across diverse and clinically-distinct populations.

Multiplexed Detection of RNA Modifications

POSTER 401

Gudrun Stengel, **Andrew Price, Zachary Miles and Byron Purse**

Gene expression patterns obtained from RNA sequencing are widely used as a surrogate for cellular protein levels. However, the correlation between the transcriptome and the proteome depends on cell state and is far from perfect. It is intriguing to assume that RNA modifications account for some of the discrepancy based on their essential roles in translation initiation and fidelity, trafficking, RNA folding and stability. Although 170 different RNA modifications have been identified today using mass spectrometry, few of the modifications can be detected accurately with sequence resolution. Most existing methods concentrate on m6A, the most prevalent modification in mRNA. To understand how RNA modifications act in concert to fine-tune the genomic message, Alida Bio has developed an assay platform with integrated bioinformatics solution that detects multiple modifications in the same reaction. The method exploits targeted, multiplexed barcoding to reveal co-localization of several modifications on the same gene, in addition to providing gene expression profiles. Alida Bio's mission is to equip translational researchers with tools that will significantly improve their understanding of gene regulation through seamless measurement of RNA modifications. Future efforts will be directed at expanding the platform to histone modifications to enable multi-epigenomics applications.

N-of-1 personalized medicine: Identification of a putative small fiber peripheral polyneuropathy- causing mutation in the glycolytic intermediate enzyme, PHGDH

POSTER 402

Max E. Glanz[1,2] , Jeanette McCarthy[3] , Tera Bowers[3,4] , Sean Hofherr[3] , Erwin Frise[3] , Martin G. Reese[3], Frank Rice[5] , Matthew E. Merritt[2] , Eric T. Wang[1]

1 University of Florida, Center for Neurogenetics, Department of Molecular Genetics and Microbiology, Gainesville, FL, USA
2 University of Florida, Department of Biochemistry and Molecular Biology, Gainesville, FL, USA
3 Fabric Genomics, Oakland, CA, USA
4 Broad Institute of MIT and Harvard, Cambridge, MA, USA
5 Integrated Tissue Dynamics (INTIDYN), Albany, NY, USA

Millions of Americans are impacted by peripheral neuropathy, a heterogenous class of neurological diseases characterized by debilitating neuropathic pain and sensory loss. One-third of peripheral neuropathies are idiopathic lacking etiological understanding, limited diagnostics, and high unmet therapeutic need. Here, we describe the characterization of a candidate gene for small fiber peripheral polyneuropathy, Phosphoglycerate dehydrogenase (PHGDH), hypothesized to be underpinning a rare, novel late onset form of PHGDH deficiency, a classical inborn error of metabolism, in a presumed n-of-1 patient. Whole genome sequencing coupled with patient medical record phenotypic analysis was performed and analyzed using Fabric GEM, an Artificial Intelligence powered platform capable of performing probabilistic disease condition matching. In this n-of-1 proband, we identified two allelic mutations in PHGDH; 1468G>A and 792+6T>G. The former reduces the catalytic activity of PHGDH while the effect of the latter is unknown. Familial genotyping analysis revealed autosomal recessive inheritance pattern with two proband children displaying 1468G>A heterozygosity, and one proband child plus the pro band's mother exhibiting 792+6T>G heterozygosity. We performed functional genetic assays to characterize the impact of the 792+6T>G variant. Mini-gene splicing assays revealed 792+6T>G results in exon 7 skipping by weakening the 5' splice site. Patient fibroblasts exhibited a reduction in PHGDH transcript and protein levels. To explore strategies for abrogating PHGDH exon 7 skipping to restore optimal enzyme levels, we carried out a high-throughput point mutagenesis screen of PHGDH exon 7 and flanking introns. This approach identified putative exon 7 regulatory splice sites serving as therapeutic targets. Given the prominent sensory nerve clinical phenotype, we sought to develop a new clinical diagnostic method capable of characterizing specific nerve fiber subtype innervation patterns in situ from 3 mm patient skin punch biopsies enabling accurate distal nerve fiber degeneration pattern identification providing an unprecedented phenotypic analysis of in situ nerve fiber death. Our work is the first of its kind to identify new functional roles for PHGDH mutants and their causal candidacy for a novel form of neuropathy. Collectively, our approach sets the stage for new genetic and clinical phenotypic diagnostics and RNA-targeting therapeutic development for polyneuropathies.

New insights into genomic variation from long-read sequencing of >1000 African-American participants in the All of Us Research Program

ABSTRACT SELECTED TALK

Kiran Garimella[1], Steve Huang[1], Yulia Mostovoy[2], Karynne Patterson[3], Matt Danzi[4], Fabio Cunial[1], Josh Smith[3], Beri Shifaw[1], Tara Dutka[5], Chelsea Berngruber[6], William Harvey[3], Sophie Schwartz[1], Emily Laplante[1], Medhat Mahmoud[7], Lee Lichtenstein[1], Yuanyuan Wang[8], Stephan Zuchner[10], Andrea Ramirez[5], Anji Musick[5], All of Us Long-Read Working Group, Kim Doheny[9], Donna Muzny[7], Michael Schatz[9], Fritz Sedlazeck10, Shawn Levy[6], Evan Eichler[3], Michael E. Talkowski[1,2]

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

The All of Us (AoU) research program is a national biobank initiative aiming to collect health and genomic data from one million participants for the purposes of advancing precision medicine. While there has been tremendous progress to date (short-read whole genome sequencing [srWGS] of >245,000 participants), recent studies have highlighted large swaths of the human genome and biomedically relevant genes that are uniquely accessible to long-read sequencing (lrWGS). Here, we present the launch of the AoU lrWGS initiative, a multi-institutional and multi-phase project to capture and openly release genomic variation from ~15,000 diverse individuals with matched srWGS and extensive phenotypic data.

The AoU long-read working group has now completed Phase 1 of the program, releasing data from 1,027 participants who self-identify as African-American or Black. We describe the novel genomic variants and human disease association discoveries only tractable from an integrated genomics dataset of this scale, which includes ~8x Pacific Biosciences lrWGS and ~30x Illumina srWGS. Our analyses to date have generated alignments on both GRCh38 and the telomere-to-telomere CHM13 reference, genome assemblies, short-variant (SNV/indel) callsets using DeepVariant, and structural variant (SV) callsets using a combination of PBSV, PAV, and Sniffles2 on all individuals.

(continue to page 78)

(continued from page 77)

These analyses have identified an average of 4-5M SNVs/small indels and ~22.4k SVs per individual. Collectively, we provide an unparalleled resource of 2.2M SVs derived from this diverse population cohort. Through downsampling analyses, we estimate current sensitivity of 80%.In our ongoing analyses, we resolve variation across repeat expansions, mobile elements, and highly complex genomic regions such as the major histocompatibility complex, LPA, BRCA1/BRCA2, etc. We further benchmark complete genome-wide characterization of the value added for SVs captured between these lrWGS and matched srWGS datasets and optimal strategies for future variant interpretation.

Complete data and results are accessible in the All of Us Researcher Workbench, as well as reproducible cloud-native workflows that can be reused in other population-scale long-read sequencing projects. Finally, we will preview Phase 2 work currently underway, which will expand long-read data generation to ~15,000 diverse participants on both PacBio and Oxford Nanopore platforms.

Novel platform for three-dimensional tumor visualization and surgical planning

POSTER 403

Amanda Parker, **Tyler Earnest, Arda Pekis, Vignesh Kannan, Evandros Kaklamanos, Joseph Peterson, Tushar Pandey, John Cole, Daniel Cook, Anuja Antony**

SimBioSys, Inc.

Introduction:
Recent results from ACOSOG Z11102 trial highlighted how improvements in medical imaging have contributed to decreasing local recurrence rates in patients with multi-ipsilateral breast cancer who de-escalate surgery and undergo lumpectomy. Currently, surgical oncologists rely on 2D medical imaging to mentally visualize 3D morphology, size, and location of tumors to determine optimal surgical options. However, this process is understandably difficult for patients, and represents a barrier to actively participating in decision-making and surgical planning. Here we present a novel platform that takes medical imaging one step further by allowing 3D tumor visualization and computing key metrics, assessing multi-focality and facilitating surgical planning.

Methods:
Utilizing DCE-MRIs as input, we developed a tumor modeling platform that generates 3D model representations of a patient's breast tissue and tumor. The platform computes useful metrics such as tumor-to-breast volume ratio, distance to surgically useful anatomical landmarks (e.g., nipple), and tumor multi-focality.

Results:
Using the 3D tumor representations, we generated labels for tumor, skin, chest wall, nipple, adipose tissue, glandular tissue, and vasculature. After tumor segmentation validation studies (n=48) were completed, the median volumetric error was 13% and the median maximum Hausdorff distance was 16.6 mm. We next determined whether the tumor is monocentric, multifocal, or diffuse, computed the tumor-to-breast volume ratio, and measured the distance of closest approach between tumor and nipple. Additionally, by adding a margin to the tumor and computing the volume within a convex boundary containing the tumor+margin, our platform
provides a pre-operative estimate of the extirpative volume and its fraction of the total breast volume.

Conclusions:
Our technology provides a surgeon the ability to visualize a tumor's location, distribution, multi-focality in the breast, and other key metrics – all in 3D. This facilitates the physician-patient conversation and can better inform and support the selection of the type of surgery.

Novel Use of Plasma Microbial Cell-Free DNA Sequencing for Detection, Quantification, Subtyping, and Genomic Assembly of Monkeypox Virus

POSTER 404

Nicholas Noll[1], Martin S Lindner[1], Kevin Brick[1], Sarah Y Park[1], Sivan Bercovici[1]

1 Karius, 975 Island Drive, Redwood City, CA 94065

In the 2022 global Monkeypox Virus (MPXV) outbreak, conventional PCR diagnostics struggled with atypical disease presentations and the emergence of divergent strains. This highlighted the need for advanced diagnostic techniques that are both precise and adaptable. We introduce a novel precision medicine approach using plasma microbial cell-free DNA (mcfDNA) sequencing for comprehensive detection, quantification, and genotyping of MPXV.

The analytical sensitivity, specificity, and precision for MPXV detection were characterized through in-silico inclusivity and exclusivity evaluations, expanding beyond our prior analytical validation studies. Fifteen plasma samples from a cohort of 12 subjects were analyzed using mcfDNA sequencing. We applied our low-coverage genotyping methodology to the MPXV mcfDNA data, identifying 22 mutations across 10 genomic loci, 21 of which were G>A/C>T transitions, mirroring the genetic trends observed during the outbreak. Phylogenetic analyses allowed for the successful classification of the isolates into distinct sublineages. Finally, the genome of MPXV was assembled de-novo, using mcfDNA data from a single patient. Outside the repetitive ITR region, the assembly resolves into one majority contig.

Crucially, this approach provides insights into the evolving pathogen landscape in the context of an outbreak, including the potential to detect acquired resistance using mcfDNA, thus augmenting disease surveillance and preparedness for emerging infectious diseases. The integration of mcfDNA sequencing into diagnostic frameworks represents a potential step forward in precision medicine. While further studies are required to fully realize its implications, initial findings suggest it could enhance patient care strategies and contribute to more robust epidemic response capabilities.

ABSTRACT SELECTED TALK

Performance of BeginNGS, an artificial intelligence-enabled genome sequencing system, for universal newborn screening, diagnosis, and precision medicine for 410 severe childhood genetic diseases in over

Meredith Wright

Rady Children's Institute for Genomic Medicine

Newborn screening (NBS) improves outcomes in selected severe childhood disorders by identification and treatment initiation for high risk infants at or before symptom onset. Expansion of the NBS Recommended Uniform Screening Panel (RUSP) has lagged leading to unnecessarily delayed diagnosis and suboptimal outcomes in hundreds of severe, early childhood onset single locus (mendelian) genetic diseases. BeginNGS supplements RUSP-NBS for these diseases by combining NBS-by genome sequencing (GS) with prioritized interpretation and virtual guidance regarding confirmatory testing, specialist referral, and urgent management. While clinical development is incomplete, here we report results of the second retrospective study of BeginNGS. Of 518 gene-disorder dyads evaluated for BeginNGS using modified Wilson and Jungner principles and Delphi methods, 410 disorders associated with 341 genes and 1,654 effective therapeutic interventions were retained, with a total incidence in newborns of ~5%. The analytic performance of BeginNGS is trained with genome sequences of large retrospective true positive (TP) and true negative (TN) cohorts. Previously we reported training of 29,865 pathogenic (P) or likely P (LP) variants associated with 388 disorders in 459,083 TP and FN individuals. Following training, BeginNGS had 3 FP/1,000 subjects and TP rate (sensitivity) of 88.8% (119 of 134 diagnoses). Here we will report results of BeginNGS version 2 with 410 disorders, ~50,000 variants from both ClinVar and Genomenon's Mastermind, as well as new artificial intelligence software for detection of novel loss of function variants (Transformer) in a larger cohort including 7,575 critically ill children with suspected genetic disorders and their parents. We will also report results of an actual/counterfactual comparison of the clinical utility, average cost per diagnosis, and incremental cost effectiveness ratio of rapid diagnostic GS ($7,000 cost per newborn with return of results during an intensive care unit admission) with BeginNGS ($500 with results on day of life (DOL) 10) and RUSP-NBS ($211 with results on DOL 10).

Prevalence and clinical significance of commonly diagnosed genetic disorders in preterm infants

POSTER 405

Selin Everett

Background: Preterm infants (<34 weeks' gestation) experience high rates of morbidity and mortality before hospital discharge. Genetic disorders substantially contribute to morbidity and mortality in related populations. The prevalence and clinical impact of genetic disorders is unknown in this population. We sought to determine the prevalence of commonly diagnosed genetic disorders in preterm infants, and to determine the association of disorders with morbidity and mortality. Methods: This retrospective multicenter cohort study included infants born from 23 to 33 weeks' gestation between 2000 and 2020. Nineteen genetic disorders were abstracted from diagnoses present in electronic health records, including aneuploidies (trisomy 13, 18, 21, Turner syndrome, Klinefelter syndromes), copy number variants (13q deletion, cri-du-chat syndrome, DiGeorge syndrome), monogenic disorders (cystic fibrosis, fragile X syndrome), and hematological disorders (sickle cell, thalassemia, hemoglobin Bart's). We excluded infants transferred from or to other healthcare facilities prior to discharge or death when analyzing clinical outcomes including death, acute kidney injury, intraventricular hemorrhage, necrotizing enterocolitis, severe bronchopulmonary dysplasia, severe retinopathy of prematurity, sepsis, and shock. We determined the adjusted odds of pre-discharge morbidity or mortality after adjusting for known risk factors including sex, gestational age, race, and birthweight. Results: Of 409,704 preterm infants, 5838 (1.4%) had genetic disorders. Of the 17,427 infants who died, 566 (3.2%) had genetic disorders. Of the 65,968 infants with a severe morbidity, 1319 (2.0%) had genetic disorders. Infants with trisomy 13, 18, 21, or cystic fibrosis had greater adjusted odds of severe morbidity or mortality. Conclusions: These findings represent the most comprehensive description of the prevalence of genetic disease and risk estimates of morbidity and mortality in preterm infants. Clinicians should consider genetic testing for preterm infants with severe morbidity and maintain a higher index of suspicion for life-threatening morbidities in preterm infants with genetic disorders. These findings can inform decision making for clinicians and families. Prospective genomic research is needed to clarify the prevalence of genetic disorders in this population, and the contribution of genetic disorders to preterm birth and subsequent morbidity and mortality.

Prioritizing variants from diverse populations to support ClinGen as a global genomic knowledge resource for scientific and clinical communities

POSTER 406

Pamela Ajuyah Robertson[1], Marina DiStefano[1], Danielle Azzariti[1], Ryan Webb[1], Kezang Tshering[1], Kathryn E. Hatchell[2], John Garcia[2], Heidi L. Rehm[1,3] on behalf of the Clinical Genome Resource

1 Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA
2 Invitae Corporation, San Francisco, CA, USA
3 Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

The Clinical Genome Resource (ClinGen) is an NIH-funded initiative developing authoritative resources to define variant pathogenicity and validity of gene-disease relationships. In the past decade ClinGen has grown globally to over 2500 members with expertise in 16 clinical domains. ClinGen currently has 52 Gene Curation Expert Panels focusing on evaluating the clinical validity of gene-disease relationships and has found that 24% of claims have insufficient evidence. This work guides the inclusion of genes in diagnostic panel testing and their interpretation in genomic analyses, highlighting the importance of ClinGen's work in healthcare. ClinGen currently supports 62 Variant Curation Expert Panels (VCEPs) developing gene and disease tailored specifications and curating variant pathogenicity. A total of 36 specifications have been released from 24 approved VCEPs and 5324 variants have been curated to date. The expertly curated variants from VCEPs are uploaded into ClinVar – an NCBI maintained database that facilitates global sharing of classified variants.

Due to the substantial number of variants per gene, VCEPs must prioritize variants that are of most benefit to the global community. To ensure curated variants are representative and inclusive of overall human genetic diversity, Invitae® contributed patient testing data documenting variation that was observed more commonly in underrepresented populations. 64 variants of uncertain significance fall under the domain of 22 ClinGen VCEPs. These variants were of interest due to functional evidence (19%), case observations (47%), segregation evidence (8%) or a combination (26%). We began extensive curation first with the Hearing Loss Variant Curation Expert Panel (HL EP). To date, 12 variants from Black or Hispanic populations have been curated and submitted with expert classifications from ClinGen. These submissions resolved 50% of variants that had conflicting classifications in ClinVar submitted by multiple labs. Evidence summaries include reference to the populations harboring these variants to explain the divergence in frequency from commonly used population databases that may not contain adequate allele frequency representation.

ClinGen's efforts to encompass broader genetic diversity in prioritizing variant curation will allow its resources to be more equitably applied to all individuals. We appreciate the generosity of Invitae in sharing testing data to enable this work.

RARe-SOURCE™: Integrated bioinformatics resource for systematic exploration of rare diseases

POSTER 407

Gregory J. Tawa[1], PhD, Uma Mudunuri[2], Daniel Watson[2], Mohammad Alodadi[2], PhD, Erica Lyons[2], PhD, Anney Che[2], PhD, Roy Satyaki[2], PhD, Richa Madan Lomash[1], PhD, London Toney[1], Cara Purdy[1], Stephanie Mounaud[1], Forbes Porter[3], MD, Sharie J. Haugabook[1], PhD, Elizabeth Ottinger[1], PhD.

1 Therapeutic Development Branch, Division of Preclinical Innovation, National Center for Advancing Translational Sciences, National Institutes of Health, 9800 Medical Center Drive, Rockville, MD, 20878
2 Advanced Biomedical Computational Science, Frederick National Laboratory for Cancer Research, Frederick, MD, 21702
3 Division of Translational Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH, Bethesda, MD, 20892.

There are thousands of rare diseases, and most have no approved treatments. Researchers, clinicians, rare disease patients, advocacy groups and other community stakeholders are marshalling resources to find therapies. The data generated from these efforts is often stored in disparate systems, each with separate rules for retrieval and use. Integration of this data into a common platform would greatly facilitate discovery of rare disease diagnostics and therapeutics. Towards this end, the National Center for Advancing Translational Sciences (NCATS) conceptualized Rare-SOURCETM and began to develop this integrated bioinformatic resource for rare diseases with Advanced Biomedical Computational Science. The goal is to establish an accessible and searchable resource that can discover commonalities in rare diseases and advance translational research. At present 125 data sources have been identified and are being integrated to create Rare-SOURCE. The data, taken from these sources and the scientific literature, spans a wide range, from molecular level details to disease consequences and potential therapies. The interface allows users to begin an inquiry starting with a gene, a disease, or both. Natural Language Processing-based Artificial Intelligence (NLP-AI) algorithms efficiently retrieve publications associated with query inputs. Deeper exploration can reveal molecular details of genetic variation and its connection to rare disease. NLP-AI algorithm workflows allow for scaling to analyze thousands of literature articles to filter those connecting rare diseases to causative variants thereby reducing the time necessary for assembly of high quality, curated, data sets for rare disease research. In summary, by connecting disparate data sources combined with AI literature mining, scientists can utilize Rare-SOURCE to conduct systematic explorations of rare disease data and accurately curate gene-disease variant associations.

Rates of genetic testing and tumor profiling in cancer patients in the United States

POSTER 408

Alexa Woodward , Pallavi Mohanty , Jean Lucia Bazan-Williamson , Ashley Brenton

1 Optum Life Sciences, Eden Prairie, MN, USA

Genetic testing and tumor profiling play a pivotal role in the treatment of cancer patients. Existing literature focuses on use of germline cancer susceptibility testing, while rates of somatic tumor profiling across demographic groups are less explored. By understanding the distribution of both germline and somatic oncology-related testing, we can identify potential disparities and target interventions to promote equitable access. This study utilized Optum's longitudinal clinical and claims databases to evaluate the utilization of genetic testing and tumor profiling among cancer patients in the United States. We included individuals with a cancer diagnosis between May 1, 2007 and February 7, 2023 (n=14,460,581). We identified patients who received molecular testing related to their cancer care (n=2,979,479). Across multiple demographic variables, we reported the proportion of cancer patients who received testing and evaluated stage at diagnosis as a potential confounder. A chi-square test of homogeneity was used to evaluate the significance of differences in proportions between demographic subgroups.

Overall, the proportion of testing varied significantly ($p < 0.0001$) among all demographic subgroups. Significantly more females received testing than males (33.1% vs 29.8%), patients with commercial insurance were most likely to receive testing, and rates of testing have generally increased year over year. Interestingly, we found higher proportions of testing among Black and Asian patients compared to White patients, perhaps because the proportion of Black patients diagnosed at stages III and IV was higher than for Whites (50.8% vs 46.4%). This analysis highlights the prevalence of molecular testing in cancer patients in the United States and its variations across demographic variables. The findings reveal some expected differences, e.g., by sex and insurance type. However, the relationship between tumor profiling and race appeared confounded by stage at diagnosis, since tumor profiling is most often used in later stage cancers and Black patients are disproportionately diagnosed at later stages. Addressing these disparities is crucial to ensure health equity in early diagnosis, testing, and personalized cancer care.

RNA transcriptome profiling in microsamples of blood

POSTER 409

Alex Chenchik[1], Lester Kobzik[1] , Mikhail Makhanov[1] , Paul Diehl[1]

1 Cellecta, Inc., Mountain View, CA, United States

Multi-omic analysis of microsamples of lancet-induced blood drops allows frequent capture and quantitation of numerous metabolites, lipids, cytokines, and proteins. Microsample-based transcriptome profiling would facilitate use of RNA biomarkers in diagnosis and treatment of immunotherapy patients, but a suitable method has not been available. We tested a targeted sequencing protocol for this purpose in human blood microsamples. We tested normal blood collected either by phlebotomy into standard vacutainer tubes or by absorption of 30 uL of blood onto a Mitra device pre-treated with an RNA-stabilization reagent. We tested detection of gene expression changes by incubating anticoagulated blood for 18 hours with endotoxin followed by isolation of identical volumes of blood using a standard method (Qiagen kit) or from Mitra devices air dried for 24 h to mimic a home-use scenario. RNA was extracted from the microsamplers. After RNA purification, a panel of 274 immune/inflammatory genes was quantified by NGS after PCR using targeted primers following the Cellecta DriverMap protocol, which avoids counting of abundant rRNA and mitochondrial sequences. The targeted sequencing results were normalized to counts per million and used to compare results in standard vs microsamples. We found robust detection of gene expression and correlation using the two methods in both unstimulated ($r = 0.94$) and endotoxin-stimulated blood. The differentially expressed genes (DEGs) identified in both standard and microsample methods showed high overlap with DEGs reported in public datasets in similar experiments. We conclude that RNA transcriptome profiling using targeted sequencing allows sensitive detection of gene expression levels in normal and activated blood samples. Because microsamples can be air-dried and mailed in for analysis, this approach has great promise for simple and repeated monitoring of RNA biomarkers in immunotherapy.

Secretome profiling captures acute changes in oxidative stress, brain homeostasis, and coagulation following short-duration spaceflight

POSTER 410

JangKeun Kim[1,2] , Nadia Houerbi[1,2] , Eliah G. Overbey[1,2] , Richa Batra[1] , Annalise Schweickart[2,3] , Laura Patras[4,5], Serena Lucotti[4,10] , Krista A. Ryon[1] , Deena Najjar[1] , Namita Damle[1] , S. Anand Narayanan[6] , Joseph W. Guarnieri[7], Gabrielle Widjaja[7] , Afshin Beheshti[8,9], Gabriel Tobias[4,11] , Fanny Vatter[4,11], Jeremy Wain Hirschberg[1] , Ashley Kleinman[1], Evan E. Afshin[1,2], Qiuying Chen[10], Dawson Miller[10], Aaron S Gajadhar[11], Lucy Williamson[11], Purvi Tandel[11], Qiu Yang[11], Jessica Chu[11], Ryan Benz[11], Asim Siddiqui[11], Daniel Hornburg[11], Steven Gross[10], Jan Krumsiek[2,3], Jaime Mateus[13] , Xiao Mao[14], Irina Matei[4,12], Christopher E. Mason[1,2,3,15,16]

1 Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY, USA.
2 The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY, USA
3 Tri-Institutional Biology and Medicine program, Weill Cornell Medicine, NY 10021, USA
4 Children's Cancer and Blood Foundation Laboratories, Departments of Pediatrics and Cell and Developmental Biology, Drukier Institute for Children's Health, Weill Cornell Medicine, New York, NY
5 Department of Molecular Biology and Biotechnology, Center of Systems Biology, Biodiversity and Bioresources, Faculty of Biology and Geology, Babes-Bolyai University, Cluj-Napoca, Romania
6 Department of Nutrition & Integrative Physiology, Florida State University, Tallahassee, FL
7 Center of Mitochondrial and Epigenomic Medicine, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA
8 Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA
9 KBR, Space Biosciences Division, NASA Ames Research Center, Moffett Field, CA, US
10 Department of Pharmacology, Weill Cornell Medicine, New York, NY
11 Seer, Inc., Redwood City, CA 94065, USA
12 Meyer Cancer Center, Weill Cornell Medicine, New York, NY 10065, USA
13 Space Exploration Technologies Corporation (SpaceX), Hawthorne, CA, USA
14 Department of Basic Sciences, Division of Biomedical Engineering Sciences (BMES), Loma Linda University Health, Loma Linda, CA 92350
15 The Feil Family Brain and Mind Research Institute, Weill Cornell Medicine, NY 10021, USA
16 WorldQuant Initiative for Quantitative Prediction, Weill Cornell Medicine, New York, NY 10021, USA

(continue to page 88)

(continued from page 87)

As spaceflight becomes more common with commercial crews, blood-based measures of crew health can guide both astronaut biomedicine and countermeasures. By profiling plasma proteins, metabolites, and extracellular vesicles/particles (EVP) from the SpaceX Inspiration4 crew, we generated "spaceflight secretome profiles," which showed coagulation, oxidative stress, and neurological pathway differences. While >93% of differentially abundant proteins (DAPs) in vesicles and metabolites recovered within six months, the majority (73%) of plasma DAPs were still perturbed postflight. Moreover, these proteomic alterations correlated better with peripheral blood mononuclear cells than whole blood, suggesting that immune cells contribute more DAPs than erythrocytes. Finally, to discern possible mechanisms leading to brain-enriched protein detection and blood-brain barrier (BBB) disruption, we examined protein changes in dissected brains of spaceflight mice. These data highlight how even short-duration spaceflight can disrupt human and murine physiology and identify spaceflight biomarkers that can help guide countermeasure development.

Semiautomated Interpretation of Genomic Disease Contribution to Unknown Causes of Infant Mortality

POSTER 411

Liana Protopsaltis[1,2], Rebecca Baer[3,4],Matthew Bainbridge[1,2],Bryant Cao[1,2], Yan Ding[1], Katarzyna Ellsworth[1,2], Laura Forero[1,2], Erwin Frise[5], Lucia Guidugli[1,2], Yong H. Kwon[1,2], Jennie Le[1,2], Eric Ontiveros[1,2], Mallory J. Owen[1], Erica Sanford[1], Lucita Van Der Kraan[1,2], Meredith Wright[1,2], Mark Yandell[5], Laura Jelliffe-Pawlowski[4], Gretchen Bandoli[3], Christina Chambers[3], Stephen F. Kingsmore[1,2]

1 Rady Children's Institute for Genomic Medicine, San Diego, CA 92123, USA.
2 Rady Children's Hospital, San Diego, CA 92123, USA.
3 Department of Pediatrics, University of California San Diego, San Diego, CA 92093, USA.
4 California Preterm Birth Initiative, University of California San Francisco, San Francisco, CA, USA.
5 Fabric Genomics, Inc., Oakland, CA 94612, USA.

Quantitative understanding of the causes of infant mortality shapes public health, surveillance, and research investments. However, the contribution of single-locus genetic diseases to infant mortality is poorly understood. We recently reported that 46 (41%) of 112 infant deaths within a large pediatric hospital system in San Diego County, CA, were associated with genetic diseases using whole genome sequencing (WGS) and semiautomated interpretation. Subsequently, we performed post-mortem WGS (>30-fold, Illumina) of de-identified, archived, newborn dried blood spots collected from 2005-2018 in an additional 914 infant deaths. Here we report initial results of diagnostic analysis of the first 211 of these infant deaths where death certificates listed ICD-10 code R95 (sudden infant death syndrome), R99 (ill-defined and unknown cause of mortality), or W75 (other accidental threats to breathing) as the cause of death. Variants identified with DRAGEN v3.9 underwent semiautomated interpretation with Fabric Genomics' Artificial Intelligence Classification Engine (ACE) and GEM with the HPO phenotype "death in infancy". Variant diplotypes in disease-associated genes with positive Bayesian GEM scores were reviewed manually according to ACMG classification guidelines. 71 (34%) of 211 infant deaths were associated with at least one pathogenic (P) or likely pathogenic (LP) variant, of which <5 were assessed to have had autosomal dominant (AD) disorders associated with infant mortality such as Left Ventricular Noncompaction (OMIM IDs 615396, 615092). The remaining 69 infant deaths either had P/LP variants in genes associated with later-onset AD disorders or carrier status for autosomal recessive (AR) disorders.

(continue to page 90)

(continued from page 89)

Analysis of variants associated with AR, mitochondrial, and X-linked disorders is ongoing and will be reported together with results of manually curated single nucleotide and copy number variants where ACE or GEM did not report. We will also evaluate this cohort for likely deleterious variants that occur recurrently and have biologic plausibility as novel causes of infant mortality. It is much more difficult to identify genetic disorders that potentially contributed to infant mortality in WGS of de-identified samples than in the previous cohort with deep phenotypes derived from electronic health records. Together these studies should clarify the contribution of single-locus genetic diseases to infant mortality in Southern California.

Sensitive identification of variants and variant phasing in newborn genetic screening using Illumina Complete Long Read (ICLR) sequencing

POSTER 412

Antariksh Tyagi[1] , Irina Tikhonova[1] , Bryan Szewczyk[1], Jane Jolicoeur[1], Evelyn Ng[1], Christopher Castaldi[1], Lauren Machado[1] , James Knight[1] , Bony De Kumar[1], Shrikant Mane[1]

1 Yale Center for Genome Analysis, Department of Genetics, Yale University, New Haven, Connecticut, USA

Accurate identification and phasing of variants are paramount for the detection of pathological genetic diseases and precision medicine. Long-read sequencing is significantly more powerful than conventional short-read sequencing in terms of its ability to detect structural variants, complex variant calling, and variant phasing. SNP phasing and sequencing of difficult regions, such as repeats and structural variants of clinical significance, are critical in newborn genetic screening. We used Illumina Complete Long Read (ICLR) to sequence DNA from newborn blood spot samples and compared it with data obtained from using PacBio long reads and Illumina SNP array. Sequencing libraries were prepared following the manufacturer's protocol and sequenced on the NovaSeq 6000 platform. The sequencing data was analyzed on the DRAGEN ICLR WGS App on the BaseSpace Sequence Hub. Sequencing metrics such as F1 scores, substitution rate, median coverage, insert size, N50 value, percentage biallelic variants, and phased heterozygous SNVs were analyzed. Additionally, the number of SNPs, non-SNPs, and structural variants obtained from ICLR was compared with PacBio SMRTTM sequencing and Illumina SNP array. The results of this study show that ICLR sequencing precision reads can provide high sensitivity for newborn genetic screening.

Short variant simulation on cfDNA sequencing to enable benchmarking performance of variant callers for liquid biopsy

POSTER 500

Anindya Bhattacharya[1], Jaspreet Kaur[1], Leo Hagmann[1] , Lucas Bishop1, Gustavo Cerqueira[1], Sudhir Chowbina[1]

1 Exact Sciences Corporation, 1812 Ashland Ave, Baltimore, MD 21205

Liquid biopsy assays measuring mutations in cell-free DNA (cfDNA) are used in cancer screening, to monitor treatment, and to track recurrence. However, cfDNA from cancer cells are often present at very low levels in the bloodstream, making mutations difficult to detect. Additionally, the mutational landscape of cfDNA can be highly variable, making it difficult to develop accurate algorithms for mutation detection.

The lack of ground truth reference sets presents challenges in liquid biopsy mutation detection algorithm development. Ground truth reference sets are known mutations that can be used to evaluate the accuracy of mutation detection algorithms.

We have developed a new short-variant simulation framework for DNA sequencing data. It is a tool written in Python to simulate reads and introduce user-defined variants, with a focus on targeted sequencing experiments.

As proof-of-principle, we applied this simulation framework to compare the performance of popular open-source variant callers: VarScan 2.0, FreeBayes, Platypus and Mutect2. Here, we simulated variant fragments with a fragment length distribution centered at 145 bp. Each variant fragment had a variant added to it. We simulated reference fragments from human reference genome hg19 with a fragment length distribution centered at 165 bp. We used 99 and 999 reference fragments to achieve a projected variant frequency of 1% and 0.1%, respectively. Each fragment represented a unique identifier (UID) family simulated polymerase chain reaction (PCR) replicates. The fragment length distributions of cfDNA were adjusted to mimic real samples.

We evaluated the performance of the variant callers for calling expected 500+ single nucleotide variants (SNVs) and insertion-deletion polymorphisms (INDELs). We compared the variant allelic frequency, depth, and strand bias of called variants to those of simulated variants.

(continue to page 93)

(continued from page 92)

Our results showed that our newly developed tool is valuable for evaluating the performance of variant callers on simulated sequencing data. We also showed that VarScan 2.0 is the most sensitive among the evaluated variant callers with >90% sensitivity.

This study provides important insights into the performance of variant callers on simulated ground truth DNA sequencing data. These insights will be valuable for the development of new liquid biopsy methods for cancer screenings and precision medicine.

Single-cell DNA sequencing reveals candidate resistant clones to third-generation tyrosine kinase inhibitors in non-small cell lung cancer

POSTER 501

Seungho Oh[1], Mina Han[2] , Chaeyeon Kim[2], Chang Gon Kim[3], Min Hee Hong[3], Hye Ryun Kim[3], and Sangwoo Kim[1,*]

1 Department of Biomedical Systems Informatics, Brain Korea 21 PLUS Project for Medical Science, Yonsei University College of Medicine, Seoul, 03722, Republic of Korea
2 Yonsei Cancer Center, Department of Internal Medicine, Yonsei University College of Medicine, Seoul, 03722, Republic of Korea
3 Yonsei Cancer Center, Division of Medical Oncology, Department of Internal Medicine, Yonsei University College of Medicine, Seoul, 03722, Republic of Korea

These authors contributed equally.

*To whom the correspondence should be addressed.

Non-small cell lung cancer (NSCLC) is a primary cause of cancer mortalities worldwide. EGFR tyrosine kinase inhibitors (TKI) are the primary therapeutic option for NSCLC with EGFR mutations. However, most patients who responded to the therapy ultimately acquire resistance with additional genetic mutations. A number of studies have reported the resistance in the third-generation TKI, but the genetic mechanisms remain poorly understood. In this study, we investigated the genetic mutations that affect the efficacy of third-generation TKI based on single-cell DNA sequencing of the tumors with acquired resistance.

Tumor tissues from four patients treated with lazertinib or osimertinib were sequenced with Illumina HiSeq2000 using Mission Bio's Tapestri® THP-V2 single-cell library preparation. We detected 237 SNVs (59.25 per sample) and 50 CNVs (12.5 per sample) using in-house and the vendor distributed pipeline. Through analysis of clonality and the burden within tumor, we identified candidate mechanisms by which resistance minor clones may survive and persist, including missense mutations in EGFR, ERBB2 and TP53, and copy number alterations in HRAS, CDK4, and STK11. Clonality analysis revealed synergistic effect of the acquired mutations and their co-existence with original driver mutations. Finally, lineage reconstruction identified the map of genetic evolution within tumors towards the acquired resistance. Our findings will provide a valuable knowledge in the arsenal to combat resistant clones to third-generation TKIs and may facilitate the discovery of new therapeutic targets in NSCLC.

Simultaneous profiling of SNP genotyping and copy number variation using 96 assays in a nanofluidic platform for pharmacogenomics studies

POSTER 502

Hui Ren , **Jian Qin , Amit Khanna , Win Hwang , Joel Brockman , Arnaldo Barican , Greg Harris , Tom Goralski , Julie Alipaz , Charles Park , David King , Naveen Ramalingam**

1 Standard BioTools Inc.
2 Tower Pl, South San Francisco, California 94080, USA

Pharmacogenomics (PGx) testing evaluates a person's genetic variation to determine how the individual may metabolize or respond to medications. It is vital in identifying responders versus non-responders to medications, optimizing drug doses, and mitigating the risk of adverse events. These genetic tests interrogate single nucleotide polymorphisms (SNP) and copy number variation (CNV) within specific genes associated with differential drug metabolism. However, the traditional qPCR-based workflow uses two separate cumbersome workflows for SNP genotyping and CNV determination. In this work, we report a unified automated workflow to simultaneously profile SNPs and CNVs in buccal swabs without extraction. The PGx panel targets 69 SNPs and three CNV targets in a single workflow using a microfluidic chip with the X9TM system. The microfluidic chip enables high-throughput sample testing against 96 assays in nanoliter volumes, which lowers the cost of reagents and reduces sample input required for the test. The PGx panel comprises 75 TaqManTM qPCR assays, including 71 genotyping assays, 3 CNV assays and 1 RPPH1 assay to serve as internal control for CNV. The 71 genotyping assays cover the most common PharmGKB Tier 1 and Tier 2 SNPs located in 17 drug response-associated genes CYP2D6, CYP1A2, CYP2B6, CYP2C19, CYP2C9, CYP3A4, CYP3A5, ABCB1, APOE, COMT, DRD2, F2, F5, MTHFR, OPRM1, SLCO1B1, and VKORC1. The 3 CNV assays in the panel cover the exon 9, intron 6, and 5' regions of the CYP2D6 gene. This PGx panel contains critical actionable pharmacogenetic variants recommended by the Clinical Pharmacogenetics Implementation Consortium (CPIC®) with clinical utility for prescribing cardiological, psychiatric, and pain management drugs. Analysis and interpretation of SNP genotyping and CNV determination is performed using a single data processing package that supports the calling of star alleles. We have assessed this panel with a total of 173 Coriell DNA samples with either known genotypes or known copy numbers. The average call rates for the SNP and CNV assays were 99.88% and 100%, and average call accuracy was 99.94% and 95.80%, respectively, for extracted genomic DNA. Three independent operators evaluated the panel on three X9TM instruments. The call rate for the extraction-free buccal swab samples was 99% and 100% for the SNP and CNV assays, respectively.

Single-Cell RNA sequencing of cerebrospinal fluid (CSF) provides insight into neurologic postacute sequelae of SARS-CoV-2 infection- a preliminary study

POSTER 503

Rose Peterson

One in 4 people are estimated to be infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) worldwide. Of these, 28% have experienced ongoing symptoms that can last months to years that can include ongoing neurologic issues such as brain fog, anxiety, depression, sleep problems, and headaches, even though the viral RNA become undetectable. It is not yet known how the infection leads to these persistent neurologic symptoms, nor what are the molecular and immunologic mechanisms associated with these conditions. Therefore, we aim to characterize the transcriptome of peripheral blood mononuclear cells (PBMC) and the paired cells isolated from cerebrospinal fluid (CSF) of individuals with neurologic postacute sequelae of SARS-CoV-2 infection (neuro-PASC) using 10X Genomics single cell RNA (scRNA) sequencing.

In this preliminary study, we included six individuals with neuro-PASC who were studied at NIH (NCT04564287), compared to five healthy volunteers. For each individual, gene expression (GEX), T cell receptor (TCR) and B cell receptor (BCR) libraries were prepared for scRNA sequencing from both PBMC and CSF cells using 10X Genomics platform. After sequencing on Illumina NovaSeq platform, these libraries were demultiplexed, and count tables were generated with Cell Ranger. Further analyses of the scRNA sequencing data were conducted using the Seurat pipeline in R.

While healthy volunteers and individuals with neuro-PASC revealed no major differences in cell type composition or cell type density in the PBMC, there were significant differences in the cell type composition and density in CSF cells of a subset of individuals with neuro-PASC. Individuals with neuro-PASC had increased regulatory T cells and plasmablasts in CSF compared to healthy volunteers. The CSF of individuals with post-covid with neurologic sequelae have TCR repertories that are unique to the individual. These results demonstrate that a compartmentalized immune profile can be observed in CSF of individuals with neuro-PASC that was not observed in peripheral blood. Additional patients are currently being evaluated.

SV and fusion gene detection by Hi-C in bulk and single cells

POSTER 504

Feng Yue1,2

1 Department of Biochemistry and Molecular Genetics, Feinberg School of Medicine Northwestern University, Chicago, IL, USA.
2 Robert H. Lurie Comprehensive Cancer Center of Northwestern University, Chicago, IL, USA.

The Hi-C technique has shown to be a promising method to detect structural variations (SVs) in human genomes. However, algorithms that can use Hi-C data for a full-range SV detection have been severely lacking. Current methods can only identify inter-chromosomal translocations and long-range intrachromosomal SVs (>1 Mb) at less-than-optimal resolution. Therefore, we develop EagleC, a framework that combines deep-learning and ensemble-learning strategies to predict a full range of SVs at high resolution. We show that EagleC can uniquely capture a set of SVs and fusion genes that are missed by whole-genome sequencing or nanopore. Furthermore, EagleC also effectively captures SVs in a panel of chromatin interaction platforms, such as HiChIP, ChIA-PET and Capture Hi-C. We apply EagleC in more than 100 cancer cell lines and primary tumors and identify a valuable set of high-quality SVs. Last, we demonstrate that EagleC can be applied to single-cell Hi-C and used to study the SV heterogeneity in primary tumors.

Targeted long-read sequencing approaches for fine-mapping genome editing events

POSTER 505

Gabriel E. Rech[1] , Chrystal Snow[2] and Chia-Lin Wei[1,2] .

1 The Jackson Laboratory for Genomic Medicine, Farmington, CT, 06032, USA.
2 The Jackson Laboratory, Bar Harbor, ME, 04609, USA.

The ability to modify genomes of model organisms using a wide range of genetic manipulation methods has provided us tools to study disease mechanisms and devise therapeutic strategies. Yet, the approaches to determine transgene integration or examine the modified loci are currently limited. The Jackson Laboratory is the world's largest source of genetically defined mice strains, sourcing more than 8,000 strains worldwide. The traditional and gold standard method for validating mouse models has been PCR based assays and Sanger sequencing. Despite its robustness, Sanger sequencing method is not highly sensitive on identifying novel low-frequency genomic alterations and can struggle to accurately characterize large and complex regions of the genomes. We have developed robust targeted long-read sequencing platforms for the identification and characterization of genome editing events in host genomes. Our established platform applies a CRISPR-Cas9 targeted approach, combined with long-read nanopore sequencing, to enhance sequencing depth at specific loci of interest. This platform provides an end-to-end workflow to fine map vector/transgene integration sites and their associated genomic aberration. The entire workflow has been optimized with a wide range of sample types. Because the robustness of the capture reactions and the simplicity of the nanopore sequencing operation, the process can be easily automated and is scalable to achieve high throughput and consistency. We expect that this platform will be valuable for vector integration assessment and reveal functional insight in the frequency, complexity, impacts and clonality of the integration events.

TCR repertoire dynamics inform durable clinical benefit in individuals on PD-L1 Treatment and can be utilized to analyze viral associations

POSTER 506

I.K. Ashok Sivakumar[1]

1 Applied Physics Laboratory, Johns Hopkins University, Laurel, MD, USA

Characterization of T-cell receptor (TCR) repertoires can contribute to understanding the adaptive immune system in the tumor microenvironment (TME) in a therapeutic setting. Here we present nuTCRacker, an open-source tool suite for characterizing and quantifying T-cell repertoires to interpret dynamic changes within and across samples. We show that NuTCRacker identifies trends across repertoires at multiple time points and across sample populations, utilizing changes in mutual information. The tool can be applied to compare T- cells in infiltrating repertoires from peripheral blood to tumor and between pre- and post-treatment samples. Using TCR sequence data from individuals in two clinical studies of PDL-1 treatment, we identified statistically significant changes in those exhibiting durable clinical benefit (DCB). Late-breaking analysis may also be included to examine precise differentiators of TCR sequences in samples diagnosed with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).

The genomic landscape of head and neck squamous cell cancer and its implication for treatment in TRIUMPH study

POSTER 507

Shinwon Hwang[1*] ,Min Hee Hong[2*],Seong Hoon Shin[3],Tak Yun[4], Keun-Wook Lee[5], Joo Hang Kim[6], Myung Ju Ahn[7], Bhumsuk Keam[8], Hyun Woo Lee[9], Min Kyoung Kim[12], Hwan Jung Yun[11], Hyo Jung Kim[12], Sang Hee Cho[13], Hyungsoon Park[14], SoYeon Oh[15], Sang-Gon Park[16], Sung-Bae Kim[17#], Sangwoo Kim[18#], Hye Ryun Kim[2] on behalf of KCSG TRIUMPH investigators.

***These authors contributed equally as the first authors**
 # These authors contributed equally as the corresponding authors.

1 Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea, Department of Medicine, Physician-Scientist Program, Yonsei University College of Medicine, Seoul, Republic of Korea
2 Divison of Medical Oncology, Department of Internal Medicine, Yonsei Cancer Center, Yonsei University College of Medicine, Seoul, Republic of Korea
3 Department of Internal Medicine, Kosin University Gospel Hospital, Busan, Republic of Korea
4 Rare Cancers Clinic, Center for Specific Organs Cancer, National Cancer Center, Gyeonggi-do, Republic of Korea
5 Seoul National University College of Medicine, Seoul, South Korea; Seoul National University Bundang Hospital, Seongnam, South Korea
6 Division of Medical Oncology, Department of Internal Medicine, CHA University Bundang Medical Center, Bundang, Republic of Korea
7 Division of Hematology-Oncology, Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
8 Department of Internal Medicine, Seoul National University Hospital, Seoul, Republic of Korea; Cancer Research Institute, Seoul National University College of Medicine, Seoul, Republic of Korea
9 Department of Hematology-Oncology, Ajou University School of Medicine, Suwon, Korea.
10 Division of Hematology-Oncology, Department of Internal Medicine, Yeungnam University Hospital, Yeungnam University College of Medicine, Daegu, Republic of Korea
11 Department of Internal Medicine, Chungnam National University Hospital, Daejeon, Republic of Korea
12 Department of Internal Medicine, Hallym University Sacred Heart Hospital, Anyang, Republic of Korea
13 Department of Hemato-Oncology, Chonnam National University Hwasun Hospital, Chonnam National University Medical School, Gwangju, Republic of Korea
14 Division of Medical Oncology, Department of Internal Medicine, St. Vincent's Hospital, College of Medicine, The Catholic University of Korea, Seoul, South Korea
15 Medical Oncology, Department of Internal Medicine, Pusan National University Yangsan Hospital, Yangsan, Republic of Korea.

(continue to page 101)

(continued from page 100)

16 Department of Hemato-oncology, Chosun University Hospital, Gwangju, Republic of Korea
17 Department of Oncology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea
18 Department of Biomedical Systems Informatics and Graduate School of Medical Science, Yonsei University College of Medicine, Seoul, Republic of Korea.

Head and neck squamous cell carcinoma (HNSCC) was ranked as the sixth most prevalent cancer globally in 2018, and its incidence continues to rise, necessitating precise genomic approaches for effective oncologic treatment. The TRIUMPH study is a phase 2 umbrella trial, comprising a multicenter, open-label, single-arm design based on genomic profiling. A total of 419 HNSCC patients were prospectively analyzed in our study to investigate the genomic landscape of HNSCC and its clinical implication. TP53, CDKN2A, PIK3CA, FAT1 and EGFR were the frequent mutations identified in our cohort, which were consistent with previous studies. We identified mutational patterns in targetable pathways and several factors that significantly influenced targeted therapy outcomes in HNSCC. Specifically, NOTCH1 mutations and MYC amplifications were associated with a poor prognosis in patients treated with therapies targeting the PIK3CA gene. Additionally, the efficacy of CDK 4/6 inhibitors was influenced by the type of CDKN2A gene mutation. We conducted a comprehensive investigation of the genomic landscape variation across different primary tumor sites, emphasizing that copy number variations were related to TP53 mutations, smoking status, and age. In the subset analysis, oral cavity cancer predominantly affected younger patients, where most TP53 mutations were identified without notable germline variants or smoking history. In HPV-related oropharyngeal cancer, we observed a distinct pattern of exclusivity between TP53 and PIK3CA single nucleotide variants or insertions/deletions. Also, HPV infection may serve as a prognostic marker for patients treated with immunotherapy.

These findings underscore the clinical implications of genomic profiling and identify potential candidates for targeted therapy in HNSCC, paving the way for more personalized and effective treatment strategies.

Key words: head and neck squamous cell cancer, genomic landscape, clinical implication

Transfer learning improves AI-guided precision cancer treatment selection

POSTER 509

Casey Sederman[1] , **Tonya Di Sera**[1] , **Gabor Marth**[1]

1 University of Utah School of Medicine, Department of Human Genetics, Salt Lake City, UT

Background: Despite a burgeoning repertoire of anticancer therapies, a lack of rational guidance for precision treatment selection remains a critical barrier to precision oncology practice. Most tumors lack identifiable, clinically actionable molecular alterations, limiting personalized treatment selection to a minority of patients. Further, tumor heterogeneity manifests in highly variable treatment response, even amongst theoretically targetable patient cohorts. Recently, deep learning (DL) models have emerged as a promising approach for precision treatment selection, achieving state of the art performance in cancer cell lines. However, when evaluated in precision oncology-relevant settings, the performance of existing models falls far below the mark. Motivated by these challenges, we developed ScreenDL, a novel DL-based cancer drug response prediction method designed for use in precision-guided therapy selection.

Methods: ScreenDL accepts a drug's chemical structure and a tumor's molecular profile as input and predicts drug response. ScreenDL was rigorously validated in cell lines and patient-derived tumor models, emphasizing the ability to predict differential drug response in never-before-seen cell lines and patients. Uniquely, ScreenDL has the ability to incorporate partial drug screening data, often obtainable in realistic clinical scenarios, through a patient-specific transfer learning strategy.

Results: Evaluation of ScreenDL in cancer cell lines shows that ScreenDL outperforms state of the art DL methods in precision oncology-relevant settings, achieving an overall Pearson correlation of 0.55 between observed and predicted response compared to 0.32 for existing methods. Incorporating partial, patient-specific drug screening data further improves overall Pearson correlation to 0.70. Importantly, biomaterial from surgical tumor resection is often available for patient-specific drug screening in practical clinical scenarios, informing an approach to precision oncology wherein deep multiomic tumor characterization and patient-specific drug screening can be integrated to provide reliable precision treatment recommendations.

Conclusion: Here, we demonstrate the utility of incorporating partial, patient-specific drug screening data with deep multiomic tumor characterization for improved cancer drug response prediction. ScreenDL represents an exciting new direction, bringing DL-based cancer drug response prediction closer to clinical application.

Translating Rapid Genome Sequencing into Clinical Testing in a Pediatric Tertiary Care Institution

POSTER 510

Jesse M. Hunter[1], Alejandro Otero Bravo[1], Ben Kelly[1], Don Corsmeier[1], Jason Garee[1], Vijayakumar Jayaraman[1], Heather Jenkins[1], Harkness Kuck[1], Grant Lammi[1], Marco Leung[1], Aimee McKinney[1], Huachun Zhong[1], Michael Rose[1], Tim Peterson[1], Spencer Ravagnani[1], Jessica Scholl[1], Eileen Stonerock[1], Theodora Matthews[1], Amy Siemon[1], Danielle Mouhlas[1], Elizabeth Varga[1], Richard Wilson[1], Elaine Mardis[1], Peter White[1], Bimal P. Chaudhari[1], Catherine Cottrell[1]

1. Nationwide Children's Hospital, Steve and Cindy Rasmussen Institute for Genomic Medicine, Columbus, Ohio, USA.

Genome sequencing (GS) is making its way into mainstream clinical genetic testing paradigms and provides a foundational resource for personalized medical care. It also provides important advantages over other testing modalities. GS dramatically reduces hands on time in the lab and requires fewer sample manipulations and PCR cycles, thus minimizing error, artifacts, and PCR bias, while providing uniform sequencing coverage. Among broad genetic testing strategies, GS demonstrates a robust diagnostic yield with potential to encompass a wide spectrum of disease-associated variation. We describe our reasoning, challenges, and process of validating and implementing clinical GS and rapid genome sequencing (RGS) at our tertiary care pediatric institution.

We implemented a clinical RGS research protocol designed to be translated into clinically validated GS and RGS tests. The research protocol was IRB approved and all work was performed in a CLIA-certified laboratory according to standardized procedures resulting in generation of 185 GS data sets used for clinical validation. We validated multiple sample types and extraction methods. Research samples were sequenced in house and simultaneously sent out for clinical rapid exome or genome sequencing at outside institutions. This provided a set of over 200 clinically reported variants that were used for variant concordance analysis for the validation. Furthermore, variant concordance, and test specificity, accuracy, and sensitivity were determined by sequencing GIAB reference samples. Mixture studies were used to determine limits of detection and demonstrated 72% SNP recall for 10% variant allele fraction across the genome at 60X coverage. We applied sequencing depth downsampling experiments to determine optimal coverage depths to maximize variant recall. We validated RGS with a preliminary report turn-around-time (TAT) of 3-5 days with a final report in 5-8 days. To meet this aggressive TAT, we optimized many processes including wetlab procedures, bioinformatics pipelines, and use of phenotype-driven variant prioritization. Furthermore, we worked to streamline and expedite hospital logistics prior to sample receipt.

(continue to page 104)

(continued from page 103)

Future work includes evaluation of additional bioinformatics pipelines including DRAGEN and additional data analysis of mitochondrial DNA, structural variants, as well as infectious disease identification.

Ultra-High Accuracy Amplicon Sequencing on ONT sequencers

POSTER 511

Christopher Vollmers[1], **Dori Z. Q. Deng**[2]

1 Department of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, California 95064, USA

2 Department of Molecular, Cellular, and Developmental Biology, University of California Santa Cruz, Santa Cruz, California 95064, USA

The sequencing of amplicons in the form of 16S genes, antibody heavy chain (IGH) transcripts, or cancer panels is a powerful tool in clinical research and diagnostics. However, Illumina sequencing technology is limited in read length and cannot sequence full-length 16S genes without complex synthetic long read approach like Loop-Seq. Amplicons theoretically within Illumina's capability - 2x300 read pairs can cover a ~550nt IGH amplicon - suffer from declining read accuracy.

Here we show that we can sequence IGH and 16S amplicons at per-base error-rates lower than 1 error in 10,000 bases (>Q40) using low-cost Oxford Nanopore Technologies sequencers. We achieved this by using a combination of concatemerization (R2C2) and unique molecular identifier (UMI) based consensus approaches. We compare this combined R2C2+UMI approach to Illumina+UMI and Illumina-based Loop-Seq approaches and find that R2C2+UMI matches or exceeds their accuracy at equal or lower per molecule cost. To make this approach easy-to-use, we have developed BD1, a highly efficient, integrated computational tool optimized for long reads, that identifies and parses UMIs from individual reads, then groups reads by UMI and generates a highly-accurate polished consensus sequences for each UMI group. R2C2+UMI in combination with BD1 represents a new gold-standard for medium-to-long amplicon sequencing, outperforming Illumina based approaches on accuracy and/or cost while being largely agnostic to amplicon length.

In 2021, the ultra-rapid Nanopore sequencing pipeline set the world record for the fastest diagnosis using dna sequencing at 7 hours 28 minutes. With further improvement to the base calling and Nanopore chemistry, we have enabled further improvements in the pipeline in accuracy, speed and cost. While leveraging the improvement in base calling accuracy with newer Guppy, the pipeline now consists of auto-scaling of the resource in the high performance cloud computing infrastructure using Ray. Auto-scaling of resources based on the rate of sequencing data generation and fine-grained control over resource-scheduling improves cost-efficiency while maintaining the speed of base-calling and alignment. Additionally, a suite of tools associate compute performance across batches of sequencing data. The pipeline can now call small variants, structural variants and CNVs within half an hour of alignment. The variant curation stage is now automated using the GEM software from Fabric Genomics. As a result, we can generate a curated list of variants within 2 hours of completing sequencing for almost 180 gigabytes of data.

Unlocking the Potential of Blended Genome Exome (BGE) for Clinical Use: Germline Testing for Polygenic Risk Scores and Rare Variant Analysis

POSTER 512

Katherine Lafferty[1], Diana Toledo[1], Marina DiStefano[1], Areesha Salman[1], Andrea Oza[1], Marissa Gildea[1], Matthew Coole[1], Matthew DeFelice[1], Jonna Grimsby[1], Samuel DeLuca[1], Candace Patterson[1], Christopher Kachulis[1], Heidi Rehm[2,1], Niall Lennon[1]

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

Technology advancements have drastically driven down the cost of genomic sequencing, yet accessibility barriers remain. Currently, clinical tools to assess risk for conditions like heart disease or cancer from germline data can include polygenic risk scores (PRS) and comprehensive analysis of disease-associated genes. While use of these tools has increased in clinics, barriers to widespread use create health disparities within underserved populations. Furthermore, some traditional methods to reduce costs for PRS, such as imputation methods from arrays, hold biases in performance for those of European ancestry, resulting in a need for more accessible and equitable solutions. One approach to combat these issues is Blended Genome-Exome (BGE) sequencing, which combines genome and exome samples together in one tube (exome library 33%, genome library 67%), generating a single CRAM file containing data from a low coverage whole genome (2-3X) and a high coverage whole exome (60-80X) on the same sample.

The clinical use of BGE for PRS and rare variant analysis holds great potential, particularly as a screening test. The ability to combine two risk assessments into one test further increases access and information for patients. We will assess this methodology to be used for PRS and rare variant detection. To validate BGE for PRS, we will leverage de-identified samples with known high risk PRS results for several complex diseases from a previously validated method: microarray, imputation, and weighted scoring. Keeping weighted scoring and number of sites used for each condition the same, BGE replaces microarray data, and GLIMPSE imputation uses high covered exome calls and low covered genome calls to create high confidence haplotypes for PRS results. For rare variants, we will analyze 20-25 samples with a range of known pathogenic or likely pathogenic variants (SNVs, small InDels, CNVs) in genes related to hereditary cancer risk to determine the sensitivity of this method. We aim to offer BGE for as low as $180 per sample. We anticipate that BGE will prove to be a robust tool with reduced cost and bias making it an ideal tool for screening the population, identifying risk for those for whom testing is otherwise inaccessible.

Untangling how cohesin folds the Drosophila genome in post-mitotic neurons

POSTER 600

Aleena Patel[1] , Kate Scuderi[2] , Alistair Boettiger[1]

1 Stanford University School of Medicine, Department of Developmental Biology, Stanford, CA, USA
2 Cornell University, Department of Molecular Biology and Genetics, Ithaca, NY, USA

The 3D structures of chromosomes packed inside the nucleus influence gene transcription and therefore function of a cell. Cohesin is a highly conserved motor protein complex that binds to DNA and applies forces to form loops and generate regions of dense contacts among distal points along the DNA polymer. More frequent DNA-DNA associations (cis-contacts) in these regions increases the probability that transcriptional enhancer elements, often not proximal to gene loci, reach their target genes. Sequestering cis-contacts spatially can also prevent regulatory non-coding sequences from associating with off-target coding genes. In the fruit fly, Drosophila melanogaster, it is unknown whether cohesin forms loops, and to what extent Drosophila genomes would change their folded states without cohesin. We use an imaging technique called Optical Reconstruction of Chromatin Architecture (ORCA) to investigate whether changes in genome folding may underlie transcriptional deregulation that would ultimately compromise tissue growth and development. ORCA traces the folded DNA polymers at a locus of interest by incorporating multiple rounds of fluorescent DNA labeling with super-resolution microscopy. As a result, we obtain polymeric representations of a snapshot of the 3D structure of a region of a chromosome in each of thousands of cells, and we can measure cis-contacts quantitatively. Here, we apply ORCA to post-mitotic neurons of the Drosophila larval brain subjected to cohesin knockdown by RNA interference, and quantify the number, size, type of loops in these cells compared to control DNA traces at several loci. ORCA measurements show that in healthy Drosophila genomes, cohesin increases the density of cis-contacts, which compacts DNA. Since neurons are post-mitotic, our results cannot be confused with cohesin's orthogonal role in segregating chromosomes during cell division. Having established a biophysical understanding of cohesin-dependent loop formation in Drosophila, we will next investigate how unfolded genomes derail gene expression in neurons.

Utility of long-read sequencing for All of Us

POSTER 601

M. Mahmoud[1,2], Y. Huang[3], K. Garimella[3], P. A. Audano[4], W. Wan [3], N. Prasad[5], R. E. Handsaker[6], S. Hall[5], A. Pionzio[5], M. C. Schatz[7], M. E. Talkowski[8,9], E. E. Eichler[10,11], S. E. Levy[12], F. J. Sedlazeck[1,2,13]

1 Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA,
2 Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA,
3 Data Sciences Platform, Broad Institute of MIT and Harvard, Cambridge, MA 02141
4 The Jackson Laboratory for Genomic Medicine, Farmington, CT 06032 USA
5 Discovery Life Sciences, Huntsville, AL 35806, USA
6 Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA
7 Department of Computer Science, Johns Hopkins University, Baltimore, Maryland, USA
8 Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02141, USA
9 Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA,
10 Genome Sci, University of Washington, Seattle, WA, USA,
11 Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA,
12 Hudson Alpha Institute for Biotechnology, Huntsville, AL 35806,
13 Department of Computer Science, Rice University, Houston, Texas, USA

The overall heritability explained by genetic studies to date is low for most complex diseases. Several variant studies have suggested associations with loci in proximity to complex genomic regions that, in aggregate, harbor ~400 medically relevant genes (e.g., SMN1, RHCE, NCF1, LPA, TPO, TNNT1, and HLA), many of which are refractory to alignment using short-read sequencing. The emergence of long-read technologies holds great promise for the detection and phasing of short (SNVs and indels) and structural variants (SVs) to disentangle the complex interactions of these mutations. Unfortunately, the error rate and high costs limit availability and application in clinical sequencing.

(continue to page 109)

(continued from page 108)

The All of Us long-read research project aims to broaden our understanding of the sequence diversity in complex genomic regions that potentially have direct implications on variation discovery in medically relevant genes, which paves the way for individualized prevention, treatment, and care. We previously identified 386 genes largely inaccessible using short reads, which are also found in genetic panels. Mutations in these genes cause different diseases such as cancer, spinal muscular atrophy, and Rh deficiency syndrome. Using these genes plus 5,027 medically relevant genes reported across studies, we investigate the utility and limitations of long-read sequencing in analyzing HapMap samples plus four conditions of two samples using different tissue and extraction methods with Illumina, PacBio HiFi, and Oxford Nanopore Technologies. In the present study, we develop a cloud workflow to analyze long-read data utilizing state-of-the-art tools for SNVs, indels, and SVs. Likewise, we identify the best combination of tools and technologies to leverage SNVs and indels precision, recall, and F-score (99.90%, 99.80%, and 99.90%). Furthermore, we highlight the utility of long-read sequencing in calling SVs in complex medically relevant genes, which enabled us to call SVs with a 0.93 F-score, exceeding Illumina 0.45. Moreover, we assess the advantages and drawbacks per technology in covering ClinVar and GIAB truth set variant identification. To conclude, given similar coverage levels, HiFi outperforms other technologies. Nevertheless, we identify several genes where ONT (SMN1, KSCNE1, MUC1, and TERT) or Illumina (KMT2C) show improved variant identification compared to HiFi.

Utilizing the Drug-Gene Interaction database (DGIdb) for precision medicine and interpretation pipelines

POSTER 602

Matthew Cannon[1], James Stevenson[1], Katie Stahl[1], Rohit Basu[1], Adam Coffman[2], Susanna Kiwala[2], Joshua F McMichael[2], Elaine Mardis, Obi L Griffith[2], Malachi Griffith[2], Alex Wagner[1,3]

1 Institute for Genomic Medicine (IGM), Nationwide Children's Hospital, Columbus, OH
2 Department of Medicine, Washington University, St. Louis, MO
3 Department of Pediatrics, The Ohio State University, Columbus, OH

Contemporary genomic medicine pipelines for clinical decision support are enabled by integration of genomic and therapeutic knowledge bases. Human curators are expected to synthesize these data to predict how patient tumors will respond to indicated therapeutics. In some pediatric cancers, these data are especially important, as the presence of a single mutation can be sufficient to predict sensitivity to a particular treatment. This time-consuming expert evaluation process is referred to as 'the interpretation bottleneck' and is one of the largest hurdles for scalable genomic medicine. The Drug-Gene Interaction database (DGIdb) addresses this bottleneck by offering tooling for identifying clinically relevant interaction and therapeutic application data to aid clinicians and researchers in interpretation and hypothesis generation, respectively. Our resource aggregates and harmonizes 70,130 gene records and 63,968 drug records from over 40 different data sources searchable through a web application. Through its free, user-friendly interface and API, DGIdb enables the identification of approved drug-gene interactions as a way to prioritize treatment decision-making. In this abstract, we present three clinical research scenarios that outline the potential of DGIdb for alleviating the interpretation bottleneck. By exploring our aggregate data using either the new HTML client (https://beta.dgidb.org) or Python package, DGIpy (https://go.osu.edu/dgipy), we uncover the potential druggability, mechanisms of action, and biological functionality for a set of mutations common to central nervous system (CNS) malignancies.

VarCat: A precision medicine platform feveloped for scalable somatic variation categorization

POSTER 603

Wesley Goar

The amount of genomics data and variant knowledge continues to vastly expand every year. Somatic profiling sequencing assays are increasingly accessible in the clinical setting due to continuous improvements in speed, data throughput, and reduced cost. However, this has only underscored the need for efficiently collecting and evaluating variant evidence, a process which remains driven primarily by human curation, and hence, limiting genomic interpretation. Variant Scientists and Clinical Laboratory Directors are tasked with evaluating ever increasing data when analyzing high-complexity molecular data, while simultaneously ensuring that data from disparate sources and knowledge bases are consistently applied to variant classification. To help address this issue, we have developed the Variation Categorizer (VarCat), a precision medicine platform designed for scalable somatic variant analysis. VarCat utilizes standards and tools from the Global Alliance for Genomics and Health and the Variant Interpretation for Cancer Consortium (VICC) to aggregate and harmonize data from many different sources that are required for the ClinGen/CGC/VICC oncogenicity guidelines and the AMP/ASCO/CAP guidelines. VarCat allows users to assess a variant using these frameworks, while also manually adding and capturing important evidence from articles used to support conclusions or any pertinent notes. The data from the assessed variant are stored in a standardized format, which can be automatically applied for review in subsequent cases, or reassessed with updated evidence, while maintaining provenance of the changes made to the assessment. Where possible, we have also built in automated code calculations from the oncogenicity guidelines, as well as automated AMP/ASCO/CAP tiering classifications on evidence from databases like CIViC or OncoKB for review and revision by variant scientists as needed. Changes made to automated classifications are captured for future analysis to determine updates needed to improve VarCat and potential refinement of community guidance on the application of variant classification frameworks. We are actively expanding the capabilities of VarCat to enhance variant interpretation workflows at Nationwide Children's Hospital and include support for VarCat use at other institutions. Furthermore, a strategic aim of VarCat is to promote the open sharing of pediatric cancer data to enhance research initiatives and clinical outcomes in this underserved population.

Variant-Level Matching, Building a Federated Platform for Rare Disease Discovery and Diagnosis

POSTER 604

Kayla Socarras[1], Corina Antonescu[3], Sean Griffith[3], Yaron Einhorn[4], Tom Bensimhon[4], Noga Yaffe[4], Netta Kabala[4], Miro Cupak[5], Heidi L. Rehm[1,2], Nara Sobreira[3]

1 Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA
2 Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA
3 Department of Genetic Medicine, Johns Hopkins University, Baltimore, MD
4 Franklin by Genoox, Palo Alto, California, USA.
5 DNAstack, Toronto, Ontario, Canada.

Over 4,800 genes have been identified as causal for Mendelian conditions, more than 1,000 of which were made using the federated platform Matchmaker Exchange (MME) (OMIM, 07/04/23, PMID: 35537081). This relied on 'two-sided' matching whereby two parties must identify a candidate gene in unsolved cases. However, a gene may not be flagged in an analyzed genome and the causal or candidate variant is never interpreted and submitted to ClinVar. In variant matching, phenotypic data is collated on individuals with the same variant to determine if the variant is causal (matching phenotype) or not (phenotype mismatch) for identifying novel disease genes or implicating variants as pathogenic in known disease genes. To facilitate this process, 'one-sided' matching is needed, where a researcher can search genomic datasets across many databases. Thus we are establishing a federated, variant-level platform for rare variants using the Beacon_v2 API (application programming interface) developed within the Global Alliance for Genomics and Health (GA4GH) and building off the foundation established by the MME and Beacon Network. Thus, a variant query can search across participating nodes where raw genomic data is stored, to return case-level phenotype information for any individual with the variant, or acknowledge that the genotype was found and provide a defined process to access more data. We started the implementation of connecting VariantMatcher and Franklin (Gennox) databases, each containing different strategies for reducing data identifiability. For example, VariantMatcher restricts search queries to 10 per day and queryable data to rare variants (below 1% allele frequency) while requiring user registration and approval. Our initial query permission across nodes will be based upon exchanging keys, as used in MME, and later use the GA4GH passport standard for data access. We are also defining consent and regulatory parameters for 'one-sided' variant matching to guide what, when, and how raw genomic data can be queried and returned. We anticipate that constructing and implementing a path to connect existing variant matching platforms will improve widespread and meaningful genomic data sharing which is much needed for improving rare disease variant interpretation and patient diagnosis.

vcfdist: Accurate benchmarking of phased variant calls in human genomes

POSTER 605

Tim Dunn[1], **Satish Narayanasamy**[1]

1University of Michigan, Computer Science and Engineering, Ann Arbor, Michigan, USA

Accurate comparison of variant calls in VCF format is critical for many applications, including clinical identification of known deleterious mutations and benchmarking the accuracy of either targeted or whole genome sequencing (WGS) pipelines. Benchmarking is notoriously difficult when multiple variants are adjacent; such variants are termed "complex" and can often be represented in multiple ways in VCF format. We show that current variant calling evaluations are biased towards certain variant representations and may misrepresent the relative performance of different variant calling pipelines. We propose possible solutions, first exploring the affine gap parameter design space for complex variant representation and then suggesting a standard. Next, we present our tool "vcfdist" and demonstrate the importance of enforcing local phasing for evaluation accuracy. We then introduce the notion of partial credit for mostly-correct calls and present an algorithm for clustering dependent variants. This approach enables us (unlike prior work) to evaluate all variants (SNPs, INDELs, STRs, and SVs) up to 10kbp in size with a consistent methodology. Lastly, we motivate using alignment distance metrics to supplement precision-recall curves for understanding variant calling performance. We evaluate the performance of all 64 phased "Truth Challenge V2" submissions and show that vcfdist greatly improves measured performance consistency across variant representations compared to prior work. Furthermore, whole-genome comparison of two VCFs on the NIST/GIAB v4.2.1 small variant benchmark completes in only 12 minutes.

Visualizing cancer mutation profiles in the National Cancer Institute's Genomic Data Commons (GDC) using ProteinPaint visualization tools: A St. Jude and NCI GDC data integration initiative.

POSTER 606

Karishma Gangwani[1], Edgar Sioson[1], Airen Zaldivar[1], Aleksandar Acić[1], Robin Paul[1], Colleen Reilly[1], Xin Zhou[1]

1 Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN, USA.

The NCI GDC is a research program that enables collaborative discovery using genomic and clinical datasets. It allows cancer researchers and bioinformaticians to search and download cancer datasets to perform high computing analysis for individual patient profiles. This enables data sharing across cancer genomic studies in lieu of precision medicine. Several technological efforts over the years have made it possible for high performance computing. One such innovative big data solution is being supplemented by our data visualization platform called ProteinPaint. ProteinPaint is a unique interactive web-based application that houses complex genomic data along with phenotypic data. The purpose of our visualization portals is to enable data exploration and re-interpretation without the need of lengthy data access approval, big data download and analysis that requires an extensive computing infrastructure or expertise. This lollipop view (https://proteinpaint.stjude.org/?genome=hg38&gene=IDH1&mds3=GDC), shows somatic coding mutations of Isocitrate dehydrogenase 1 (IDH1) in all open-access cases. IDH1 co-mutational signature in tumor types such as gliomas can be further visualized in the OncoMatrix plot. Genome-wide mutation patterns of individual patients can be visualized using the disco plot which can be accessed interactively through the OncoMatrix plot or the Lollipop chart. Filters can be applied to narrow down cases based on demographics such as age, gender, or diagnoses parameters such as age at diagnosis, primary diagnosis etc.
A user-centric demo of ProteinPaint can be viewed at https://proteinpaint.stjude.org. Users may choose to visualize currently available open access data or upload their own datasets for visualization. Our visualizations such as Lollipop (coding mutations and gene fusions), OncoMatrix, and BAM track are currently being integrated into the GDC portal with access to controlled data and will be available late 2023.

T-cell and B-cell receptor repertoire profiling for biomarker discovery

POSTER 609

Alex Chenchik[1] , Mikhail Makhanov[1] , Tianbing Liu[1] , Dongfang Hu[1] , Paul Diehl[1] , and Lester Kobzik[1]

1 Cellecta, Inc., Mountain View, CA, United States

T-cell receptor (TCR) and B-cell receptor (BCR) repertoire profiling, also referred to as adaptive immune receptor repertoire (AIRR) profiling, holds great potential for the understanding of disease mechanisms and for the development of new treatments in infectious disease, autoimmunity and immuno-oncology. This potential could be greatly improved by combining information about receptor clonotypes with immunophenotypes of T- and B-cells. We developed a new technology for combined profiling of all human TCR and BCR variable regions and phenotypic characterization of immune cells in the same workflow. The TCR and BCR immunophenotyping method involves RT-PCR amplification and sequencing of the CDR3 regions of the TCR and BCR genes, as well as determining the expression levels of the most informative T- and B-cell phenotyping genes. Results show that this method allows for comprehensive profiling of all seven TCR and BCR chains from a single sample, in a highly reproducible manner, directly from micro-samples including cancer tissue, whole blood, sorted cells and more. Bioinformatic analysis of the next-generation sequencing (NGS) data allows profiling of the TCR and BCR clonotypes and identification of major immune cell subtypes and their activation status. Preliminary data from rheumatoid arthritis blood samples and others will be presented.

Index

	Last Name	First Name	Organization / Company	Page(s)
H	Hogg	Grant	Labcorp Oncology	50, 51
	Hou	Ying-Chen Claire	The Steve and Cindy Rasmussen Institute for Genomic Medicine	41, 42
	Hu	Daniel	UCLA	63
	Hunter	Jesse	Nationwide Children's Hospital	103, 104
	Hwang	Shinwon	Yonsei University	100, 101
J	Jovanovich	Stevan	S2 Genomics, Inc.	29
	Juarez	Edwin	Rady Children's Institute for Genomic Medicine	28
K	Kang	Seungseok	Yonsei University College of Medicine	24
	Kim	JangKeun	Weill Cornell Medicine	87, 88
	Knuckles	Philip	Roche	56
	Kolmogorov	Mikhail	National Cancer Institute, NIH	67, 68
L	Lafferty	Katherine	The Broad Institute of MIT and Harvard	106
	Li	Lewyn	Assembly Biosciences	18
	Li	Qiuhui	Johns Hopkins University	37
	Liu	Pingfang	New England Biolabs Inc	19
	Looney	Timothy	Singular Genomics	30
M	Mahmoud	Medhat	Human Genome Sequencing Center, Baylor College of Medicine	108, 109
	Maier	Keith	EpiCypher Inc.	27
	Mallory	Benjamin	University of Washington	52
	Mastoras	Mira	UC Santa Cruz Genomics Institute	58
N	Negi	Shloka	University of California, Santa Cruz	22, 23
	Noll	Nick	Karius	80
O	Oh	Seungho	Yonsei University College of Medicine	94
	Olsen	Lauren	Rady Children's Institute for Genomic Medicine	20, 21
P	Parker	Amanda	SimBioSys, Inc.	79
	Patel	Aleena	Stanford University	107
	Peterson	Rose	National Institutes of Health	96
	Protopsaltis	Liana	Rady Children's Institute for Genomic Medicine	89, 90

	Last Name	First Name	Organization / Company	Page(s)
R	Ramalingam	Naveen	Standard BioTools	95
	Rech	Gabriel E.	The Jackson Laboratory	98
	Robertson	Pamela	Broad Institute of MIT and Harvard	83
S	Sarangi	Srikant	Paradigm4	25
	Sederman	Casey	University of Utah	102
	Shah	Shreyas	Revvity	54
	Shi	Zonggao	St Jude Children's Research Hospital	53
	Sivakumar	Ashok	Johns Hopkins University Applied Physics Laboratory	99
	Socarras	Kayla	Broad Institute of Harvard and MIT	112
	Stengel	Gudrun	Alida Biosciences Inc.	75
	Stergachis	Andrew	Division of Medical Genetics, Department of Medicine, University of Washington	43
	Sun	Jiayi	MyOme, Inc.	40
T	Tang	Paul	AccuraGen	46
	Tawa	Gregory	NIK National Center for Advancing Translational Sciences	84
	Tejura	Malvika	Genomic Sciences Department, University of Washington	66
	Tshering	Kezang C.	The Broad Institute of MIT and Harvard	55
	Tyagi	Antariksh	Yale University	91
V	Vollmers	Christopher	UC Santa Cruz	105
	von Alvensleben	Giacomo	EPFL	47
W	Wan	Chen	Genome BC	57
	Ward	Alistair	University of Utah	35, 36
	Wee	Kathleen	Canada's Michael Smith Genome Sciences Centre	61, 62
	Weisburd	Ben	Broad Institute	48, 49
	Woodward	Alexa	Optum Life Sciences	85
	Wright	Meredith	Rady Children's Institute for Genomic Medicine	81
Y	Yue	Feng	Northwestern University	97