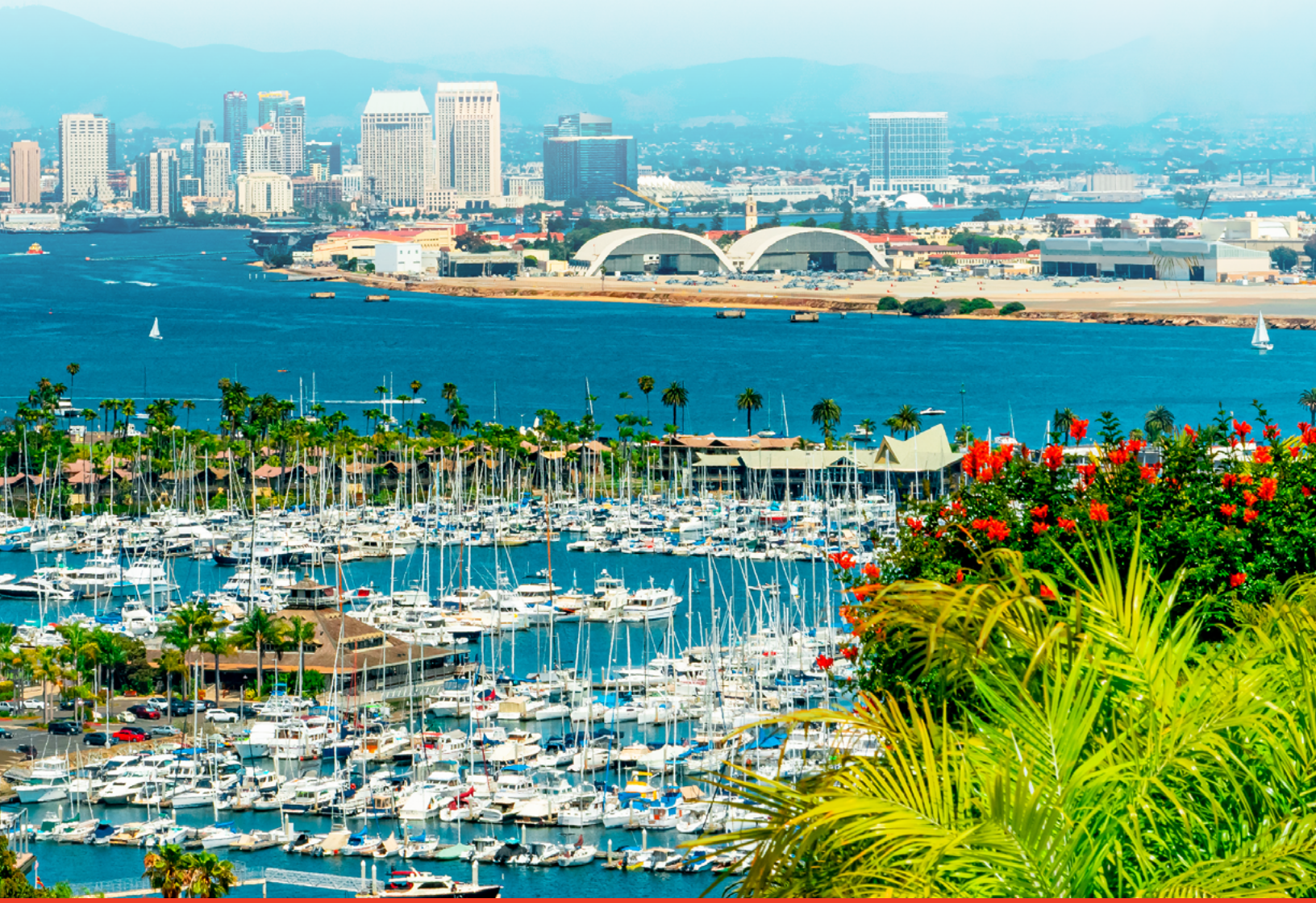
AGBT2022™

PRECISION HEALTH

September 8 - 10, 2022 • San Diego, CA



HILTON SAN DIEGO BAYFRONT

AGBT 2022™
PRECISION HEALTH

September 8 - 10, 2022 • Hilton Bay Front, San Diego, CA

It is our honor to welcome you to the 2022 Advances in Genome Biology and Technology (AGBT) Precision Health meeting.
Now in its 7th 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 6-9, 2023) and AGBT Agriculture (March 27-29, 2023).

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 22 a content rich experience.

Penelope Bonnen
Baylor College of Medicine

Wendy Chung
Columbia University Medical Center

Eric Green
National Human Genome Research Institute

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

Professor Dame Sue Hill
NHS, England

Stephen Kingsmore
Rady Children's Institute for Genomic Medicine

Howard McLeod
Geriatric Oncology Consortium, Intermountain Precision Genomics

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

AGBT2022™ PRECISION HEALTH

We gratefully acknowledge the generous support provided by the following companies

MAJOR SPONSORS

DNASCRIPT

FABRIC GENOMICS

cambridge consultants Part of Capgemini Invent

BECKMAN COULTER Life Sciences

CONTRIBUTING SPONSORS AND EXHIBITORS

ThermoFisher SCIENTIFIC

SOPHiA GENETICS™

molecular loop

SINGULAR GENOMICS

Intermountain Healthcare

L7INFORMATICS

GRAIL

GENOME INSIGHT

BLUE CAT BIO

NEW ENGLAND BioLabs Inc.

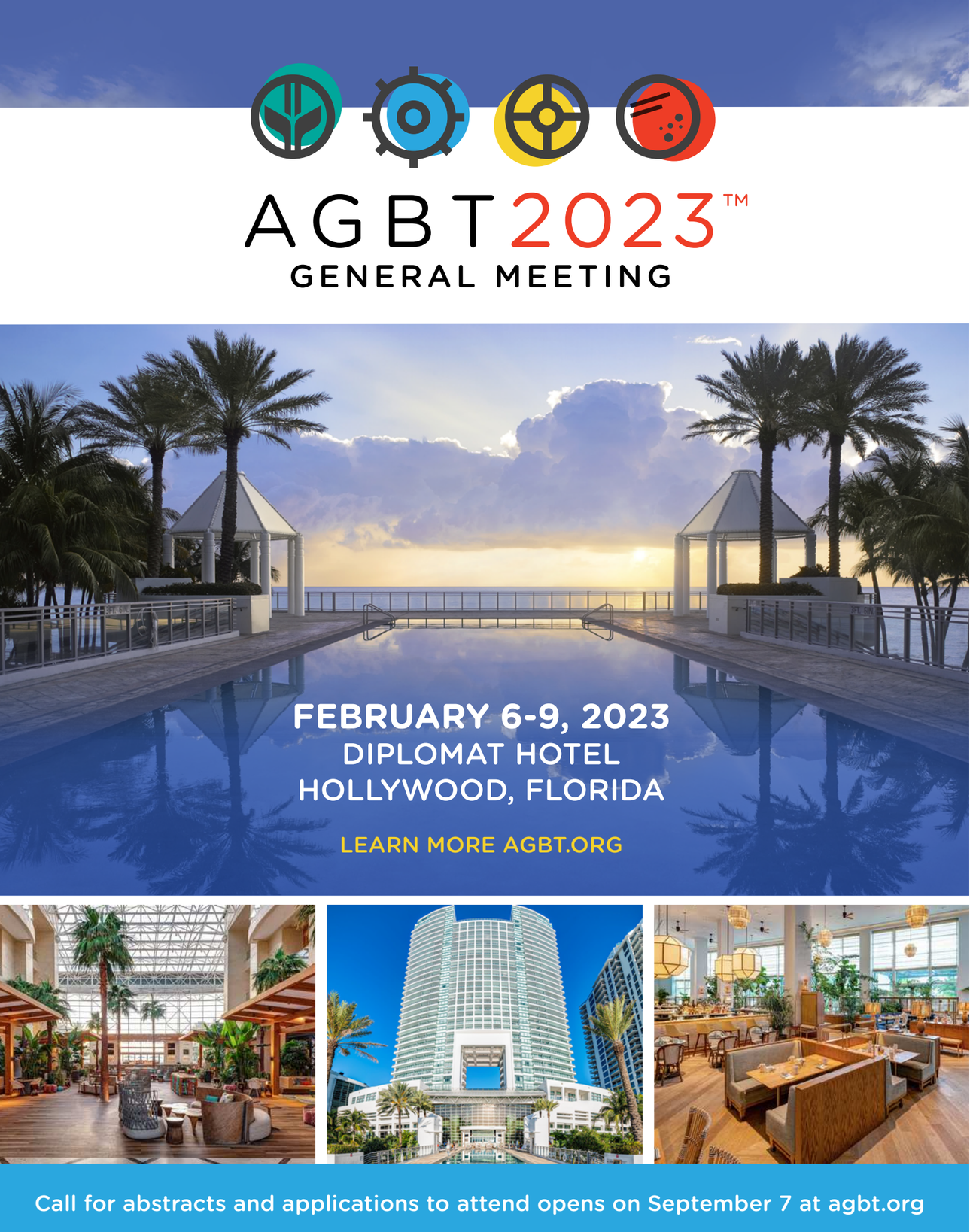
Element Biosciences

ARIMA GENOMICS

INSIDE PRECISION MEDICINE

WATCHMAKER GENOMICS

SAMPLED



General Information

Hospitality Desk

Thursday,	September 8, 2022	1:00pm – 7:30pm
Friday,	September 9, 2022	8:00am – 6:00pm
Saturday,	September 10, 2022	8:00am – 6:00pm

Name Badges

Please wear your name badge at all times. Security will refuse entry to anyone not associated with the AGBT meeting.

Transportation

San Diego International Airport is 5.8 miles from the Hilton San Diego Bayfront. The transfer time to the hotel is approximately 15 minutes. Airport transportation to and from the resort can be arranged by visiting **San Diego International Airport.**

Social Events

AGBT Welcome Dinner – **Thursday, September 8, 7:15pm – 10:00pm,** Promenade Plaza
AGBT Dinner – **Friday, September 9, 6:30pm – 9:00pm,** 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.

Notice of Vaccination Requirement for Meeting Attendees

AGBT will require proof of COVID-19 vaccination for all attendees, meaning all doses and boosters outlined by the CDC. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/stay-up-to-date.html#recommendations AND a negative PCR or antigen test taken within 72 hours of the meeting. Attendees who do not have proof of the above will not be allowed in, and registration will not be refunded.
AGBT strongly recommends masks while indoors.

SUM2Win

Make sure to visit the participating exhibitors to compete in our SUM2Win challenge! Attendees will gather numbers from each exhibitor to then calculate the final SUM. Those who have the correct final SUM will be entered to win a (1) complimentary 3-day registration to AGBT PH 2023 or (1) one year Wine Subscription! The prizes will be awarded during Saturday's closing session. (This prize is non-transferable.). We look forward 2 having SUM fun with you!

Start Tweeting!

#AGBTPH

Agenda

Thursday, September 8 (Evening)

1:00 p.m. – 7:30 p.m.	**Hospitality Desk** • Indigo West Foyer
5:00 p.m. – 5:10 p.m.	**Welcome**
Session I:	**New Technologies and Broad Perspectives in Genomic Medicine (Eric Green, National Human Genome Research Institute and Michael Talkowski, Massachusetts General Hospital and Broad Institute Chairs)** • Indigo Ballroom CDGH
5:10 p.m. – 5:45 p.m.	**Karen Miga, University of California, Santa Cruz** *"Advancing long-read de novo genome assembly methods in clinical research"*
5:45 p.m. – 6:20 p.m.	**Diana Bianchi, National Institute of Child Health and Human Development (NICHD) and National Human Genome Research Institute (NHGRI)** *"Prenatal precision medicine: The earlier the better for mom and baby"*
6:20 p.m. – 6:55 p.m.	**Daniel MacArthur, Center for Population Genomics, Garvan Institute of Medical Research** *"Towards equitable genomic medicine"*
6:55 p.m. – 7:15 p.m.	**Emily Mrig (Abstract Selected) University of California San Francisco** *"When covering and discovering are at odds: How the logic of health insurance undercuts the promise of precision medicine for people with hereditary cancer risks"*
7:15 p.m. – 10:00 p.m.	**Welcome Reception** • Promenade Plaza

Friday, September 9, 2022 (Morning)

7:30 a.m. – 9:00 a.m.	**Breakfast** • Indigo Terrace
8:00 a.m. – 6:00 p.m.	**Hospitality Desk** • Indigo West Foyer
Session II:	**Genomic Medicine I: Prenatal Diagnostics (Wendy Chung, Columbia University and Diana Bianchi, NICHD, Chairs)** • Indigo Ballroom CDGH
9:00 a.m. – 9:25 a.m.	**Ronald Wapner, Columbia University Irving Medical Center** *"Fetal sequencing of congenital anomalies: The need for clinical laboratory collaboration"*
9:25 a.m. – 9:50 a.m.	**Hamish Scott, University of South Australia, Centre for Cancer Biology** *"Slow, slow, quick, quick, slow: Genomic testing and research follow-up turnaround times affect both clinical yield and utility in genomic autopsy of pregnancy loss and perinatal death"*
9:50 a.m. – 10:15 a.m.	**Gretchen Bandoli, University of California, San Diego** *"Advancing scientific approaches through administrative data"*
10:15 a.m. – 10:35 a.m.	**Harrison Brand (Abstract Selected) Massachusetts General Hospital** *"A comprehensive and non-invasive fetal screen of all coding variation"*
10:35 a.m. – 11:05 a.m.	**Coffee Break/Exhibit Hall** • Indigo Ballroom ABEF
Session III:	**Panel Session: Economic Barriers and Predictive Value to Genome Sequencing in Precision Medicine (Kathryn Phillips, University of California San Francisco, Chair)** • Indigo Ballroom CDGH
11:05 a.m. – 12 p.m.	**Ryan Taft, Illumina; Keith Yamamoto, UCSF; Sharon Terry, Genetic Alliance**

Friday, September 9, 2022 (Afternoon)

12:00 p.m. – 1:10 p.m.	**AGBT Lunch** • Indigo Terrace
1:15 p.m. – 2:00 p.m.	**Sponsor Workshops** • Indigo Ballroom CDGH
1:15 p.m. – 1:40 p.m.	**Diamond Sponsor – DNA Script** **Chris Barbazette and John Lesnick** *"Develop genetic testing assays faster by printing the needed oligos at your benchtop"*
1:40 p.m. – 2:00 p.m.	**Emerald Sponsor – FABRIC Genomics** **Heidi Rehm, Ph.D. – Institute Member, Co-director of the Program in Medical and Population Genetics** **Diana M. Toledo, Ph.D., MS, CGD – Associate Lab Director – Clinical Research Sequencing Platform, Broad Genomics** *"A validation study of whole genome sequencing using Fabric Genomics AI-based Platforms ACE and GEM at the Broad Institute's Clinical Laboratory"*
Session IV:	**Genomic Medicine II – Rare Disease and Public Health Initiatives (Professor Dame Sue Hill, NHS England and Howard McLeod, Precision Health, Intermountain Healthcare, Chairs)** • Indigo Ballroom CDGH

Agenda

2:00 p.m. – 2:30 p.m. **Euan Ashley, Stanford University**
"The whole genome in precision health"

2:30 p.m. – 3:00 p.m. **Sharon Terry, Genetic Alliance**
"iHope genetic health: diagnosing the world's undiagnosed"

3:00 p.m. – 3:20 p.m. **Bruce Korf (Abstract Selected) University of Alabama, Birmingham**
"The Alabama Genomic Health Initiative: Five Years In"

3:20 p.m. – 5:20 p.m. **Poster Session & Exhibit Hall, Wine and Cheese Reception**
• Indigo Ballroom ABEF

6:30 p.m. – 9:00 p.m. **AGBT Dinner**
• Promenade Plaza

Saturday, September 10, 2022 (Morning)

7:30 a.m. – 9:00 a.m. **Breakfast**
• Indigo Terrace

8:00 a.m. – 6:00 p.m. **Hospitality Desk**
• Indigo West Foyer

Session V: **Advances in Genome Biology (Heidi Rehm, Massachusetts General Hospital and Broad Institute and Len Pennacchio, Lawrence Berkeley National Laboratory, DOE Joint Genome Institute, Chairs)**
• Indigo Ballroom CDGH

9:00 a.m. – 9:30 a.m. **Kelly Frazer, University of California San Diego**

9:30 a.m. – 10:00 a.m. **Stephen Montgomery, Stanford University**
"Interpreting rare variants in common and rare genetic diseases"

10:00 a.m. – 10:20 a.m. **Caralynn Wilczewski (Abstract Selected) National Human Genome Research Institute**
"Genotype first: Clinical genomics research through a reverse phenotyping approach"

10:20 a.m. – 10:55 a.m. **Coffee Break & Exhibit Hall**
• Indigo Ballroom ABEF

11:00 a.m. – 12:00 p.m. **Fireside Chat: Patient stories (Stephen Kingsmore, Rady Children's Hospital, Chair)**
Indigo Ballroom CDGH

Nathaly Sweeney, Assistant Professor of Pediatrics, Neonatal-Perinatal Medicine, University of San Diego, and Rady Children's Hospital
Jerica Lenberg, Genetic Counselor, Rady Children's Institute for Genomic Medicine
The Wingate Family

12:00 p.m. – 1:00 p.m. **AGBT Lunch**
• Indigo Terrace

1:00 p.m. – 1:30 p.m. **Sponsor Workshops**
• Indigo Ballroom CDGH

1:03 p.m. –1:18 p.m. **Sapphire Sponsor – Cambridge Consultants**
Dr. Symon Cotton, head of life sciences, Cambridge Consultants
"Sequencing 2030 – the challenges and opportunities"

1:18 p.m. – 1:30 p.m. **Ruby Sponsor – Beckman Coulter Life Sciences**
Calvin Cortes, M.P.H.,M.B.A.
"Optimizing your NGS lab with a game-changing, easy-to-use automation solution"

Session VI: **National Initiatives in Genomic Medicine: All of Us and Genomics England (Geoffrey Ginsburg, All of Us Research Program, NIH and Professor Dame Sue Hill, NHS England, Chairs)**
• Indigo Ballroom CDGH

1:30: p.m. – 1:40 p.m. **Overviews from the Chairs**

1:40 p.m. – 2:35 p.m. **The All of Us Research Program**

1:40 p.m. - 1:55 p.m. **Anji Musick, All of Us Research Program**
"Progress with all the All of Us research program 2022"

1:55 p.m. – 2:15 p.m. **Eric Venner, Baylor College of Medicine**
"The frequency of pathogenic variation in the All of Us cohort reveals ancestry-driven disparities"

2:15 p.m. – 2:35 p.m. **Niall Lennon, Broad Institute of MIT and Harvard**
"*Taking the long view: driving discovery and advancing genomic medicine with new data types in AoU"*

2:35 p.m. -3:25 p.m. **Genomics England**

2:35 p.m. – 3:00 p.m. **Richard Scott, Chief Medical Officer, Genomics England**
"The UK newborn genomes programme"

3:00 p.m. -3:25 p.m. **Augusto Rendon, Chief Bioinformatician, Genomics England**
"UKs 100,000 genomes program update and beyond"

3:25 p.m. – 3:45 p.m. **Coffee Break**
• Indigo Foyer AB

Agenda

Session VII:	**Neurodevelopmental Genomics (Penelope Bonnen, Baylor University)** • Indigo Ballroom CDGH
3:45 p.m. – 4:15 p.m.	**Stephan Sanders, University of California, San Francisco** *"Genes, isoforms, and function in neurodevelopmental disorders"*
4:15 p.m. – 4:45 p.m.	**Kaitlin Samocha, Massachusetts General Hospital** *"Investigating the contribution of rare variation to severe developmental disorders"*
4:45 p.m. - 5:00 p.m.	**Final Comments and SUM2Win Prize Winners Announced (Stephen Kingsmore, Rady Children's Hospital)**
	Dinner on your own

Posters

Accurate classification of intrapulmonary metastasis and multiple primary lung cancer from multi-scale genomic profiles

POSTER 105

Jeongsoo Won[1,†], **Yeon Seung Chung**[2,†], **Se-Young Jo**[1], **Jiho Park**[1], **Hyo Sup Shim**[2,*], **and Sangwoo Kim**[1,*]

1 Department of Biomedical Systems Informatics, Graduate School of Medical Science, Brain Korea 21 Project, Yonsei University College of Medicine, Seoul 03722, South Korea
2. Department of Pathology, Severance Hospital, Yonsei University College of Medicine, Seoul 03722, South Korea

Lung cancer often exhibits more than one tumor in multiple lesions at diagnosis, which can be either independently developed primary tumors (multiple primary lung cancer; MPLC) or metastatic tumors from one primary (intrapulmonary metastasis; IPM). While accurate discrimination of the two multifocal tumors is clinically important, the robustness of current rule-based strategies is uncertain, due to the potential dependency on the applied sequencing platforms. Here, we developed a new Bayesian probabilistic model MeTel (Metastasis-Teller for lung cancer) that classifies MPLC and IPM. MeTel computes and compares normalized likelihoods of MPLC and IPM in the given somatic mutation profiles, with considering the size of sequencing panels, variant allele frequencies and background occurrence of the mutations to enable robust and unbiased classification. Testing MeTel on 290 publicly available samples (196 MPLC and 94 IPM) sequenced in various sizes of panels (22-605 genes) from six cohorts achieved a 97% accuracy, outperforming previous methods (87-96%). We also confirmed the consistent accuracy of MeTel in larger sequencing platforms (523 genes to whole exome) in 13 in-house samples (3 MPLC and 10 IPM). Adding histological and clinical information further increased the accuracy of MeTel to achieve near-perfect classification. Our results verify the accuracy of sequencing-based classification of MPLC and IPM, and justify its clinical utility in diagnosing multifocal lung cancer.

Accurate short variant calling from sequencing data with pangenomes and DeepVariant

POSTER 110

Jean Monlong[1], **Adam M. Novak**[1], **Charles Markello**[1], **Pi-Chuan Chang**[2], **Andrew Carroll**[2], **Benedict Paten**[1]

1 UC Santa Cruz Genomics Institute, Santa Cruz, CA, USA
2 Google Inc, Mountain View, CA, USA

Accurate identification of genomic variants is critical for genetic studies and, more and more, for clinical genetic testing. With genome sequencing technology, the whole-genome can be assessed and a comprehensive spectrum of genomic variants could in theory be called. Some genomics regions and variants are in practice still challenging to detect with the standard techniques, which is to first align the sequenced reads to a reference genome. A new approach has recently shown better results by a reference pangenome instead, i.e. a reference representing multiple genomes and incorporating information about known variants a priori. At the forefront of human pangenomics, the Human Pangenome Reference Consortium (HPRC) aims at producing high-quality phased de novo assemblies for more than 300 diverse individuals, and build a comprehensive and representative human pangenome. With its draft representing 45 individuals, we show that using the HPRC pangenome improves the mapping of short reads and, as a result, the identification of short variants in an individual. Our approach maps the short Illumina reads to the pangenome using vg Giraffe and calls short variants with DeepVariant using new models trained on the pangenome. We compared our approach to the standard technique of mapping reads to the linear reference genome with BWA-MEM, or the more recent Dragen pipeline. Using both the Genome in a Bottle and the Challenging Medically-Relevant Genes truthsets, we observed a better performance for our pangenomic approach. It makes about 34% less errors and is capable of calling short variants in a larger fraction of the genome, including in more challenging regions. As expected, the calling accuracy increases when trio sequencing data is available and is still superior when mapping the short-reads to the HPRC pangenome rather than to the linear reference genome or using the DragenGRAPH mapper. In addition to quantifying the benefit of this pangenomic approach, we share the resources and workflows to accurately call small variants from sequencing data using the HPRC draft pangenome.

Addressing the direct-to-consumer genetic testing knowledge gap for non-genetics healthcare professionals

POSTER 115

Inter-Society Coordinating Committee for Practitioner Education in Genomics (ISCC-PEG) DTC-GT Working Group

Nguyen Park[1], Kathy Blazer[2], Rachel Mills[3], Kathyrn Garber[4], Heewon Lee[5], Roseann Gammal[6], Linda Ho[7], Katherine Hyland[8], Kimberly Jacoby Morris[9], Mylynda Massart[10], Elena Flowers[8], Sarah Robbins[9], Grace Kuo[11], Dyanna Christopher[9], Christopher Gunter[9], Matthew Taylor[12], Donna Messersmith[9], Tracey Weiler[13], Houriya Ayoubieh[14]

1 American College of Medical Genetics and Genomics/National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA
2 City of Hope Comprehensive Cancer Center, Duarte, CA, USA
3 University of North Carolina, Greensboro, NC, USA
4 Emory University School of Medicine, Atlanta, GA, USA
5 Park Nicollet Frauenshuh Cancer Center, Saint Louis Park, MN, USA
6 Massachusetts College of Pharmacy and Health Sciences, Boston, MA, USA
7 National Center for Advancing Translational Science, Bethesda, MD, USA
8 University of California, San Francisco, CA, USA
9 National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA
10 University of Pittsburgh, Pittsburgh, PA, USA
11 Oregon State University, Portland, OR, USA
12 University of Colorado, Denver, CO, USA
13 Florida International University, Miami, FL, USA
14 Texas Tech Health Sciences El Paso, TX, USA

Direct-to-consumer genetic testing (DTC-GT) is a convenient method for people to obtain genetic information. Over nine million U.S. individuals have used DTC-GT services. Healthcare providers are encountering patients with questions regarding DTC-GT. Of surveyed Kaiser Permanente physicians, 35% reported patients had shared DTC-GT results in the past year. In a recent survey of physician assistants (PAs), over 37% of their patients had questions regarding genetics/genomics including regarding DTC-GT, and ~30% had ordered genetic testing for their patients in the past year. Despite the growing need to manage these questions, only 12% of surveyed primary care physicians agree that physicians have sufficient knowledge to help patients understand DTC-GT results. Of surveyed PAs, only 5% felt either very prepared or extremely prepared to do so. Of surveyed consumers who shared DTC-GT results with their doctors, 27% disagreed with the idea that their physician understood genetics enough to advise them on the implications of their DTC-GT results. Due to this perceived lack of knowledge, many healthcare professionals are reluctant to tackle genomic medicine in their clinical practice and would seek opportunities to enhance their genomic knowledge. The National Human Genome Research Institute's Inter-Society Coordinating Committee for Practitioner Education in Genomics (ISCC-PEG) DTC-GT Project Group consists of a spectrum of genetics professionals who aim to improve the genomic literacy of healthcare professionals and

enhance the effective practice of clinical genomic medicine around DTC-GT. The ISCC-PEG DTC-GT Project Group created a DTC-GT for Healthcare Professionals Frequently Asked Questions (FAQ) resource housed on Genome.gov. The FAQ addresses benefits and limitations of DTC-GT, differences between DTC-GT and clinic-based testing, explores types of results that DTC-GT may produce, including carrier-testing, reviews pharmacogenomic DTC-GT testing validity, and provides recommendations to healthcare providers regarding clinical decision-making related to DTC-GT results. In addition to the FAQ, the DTC-GT Project Group participated in NHGRI Healthcare Provider Genomics Education Week, presenting an educational webinar and short YouTube videos about DTC-GT. The Project Group is also developing a point-of-care tool for healthcare professionals to facilitate an outpatient encounter pertaining to DTC-GT with 11 DTC-GT patient scenarios used to validate it. The point-of-care tool has been implemented in the Qualtrics software, as a proof of concept, with the intention to develop a mobile app/website that can support healthcare professionals as they respond to their patients' queries about DTC-GT. Products and outcomes from the DTC-GT Project Group's work will be presented.

Azurify: Pathogenicity classification of somatic mutations in precision medicine through online machine learning

POSTER 125

Ashkan Bigdeli[1,2], **Robert B Faryabi**[2,3]

1 Drexel University, School of Biomedical Science, Philadelphia, PA, USA
2 Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, USA
3 Penn Epigenetics Institute, University of Pennsylvania, Philadelphia, PA, USA; Abramson Family Cancer Research Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

The accurate classification of somatic mutations using next-generation sequencing (NGS) data has become integral to diagnostics and prognostics pan-cancer; yet the classification of these mutations remains highly manual, inherently variable, and largely inaccessible outside of specialized laboratories. Here we introduce Azurify, a method that leverages the power of online machine learning, specifically gradient-boosted decisions trees, to operate on a feature set derived from AMP/ASCO recommended resources to determine the pathogenicity of mutations produced by DNA NGS. Azurify was trained using over 479,500 clinically classified mutations by way of 8,202 patients and 147 cancer phenotypes. Azurify achieves 98.81% accuracy on disease associated mutations, 95.63% accuracy in variants of unknown significance, and 99.89 % accuracy in benign variations when tested on holdout data containing single-nucleotide and small insertion/deletion events typically seen in clinical NGS assays. Azurify not only increases the accessibility of clinical mutation classification by being distributable but provides a framework for the accurate automation of intra-lab analysis that could impact the turnaround time of lab ordered somatic testing. Lastly, Azurify prompts the discussion on the potential benefits of inter-lab model training as a means of decreasing the known variabilities in somatic mutation classification.

Blood metagenomic sequencing in young children detects DNA signatures of pathogenic and commensal microbes that correlate with clinical diagnosis and immune repertoire

POSTER 130

Prithvi Raj[1,4], **Tulasi Relangi**[1,4], **Bo Zhang**[1,4], **Carlos Arana**[1,4], **Nancy Kelly**[2,5], **Lora Hooper**[1], **Nicolai S.C. van Oers**[1,3]

1 Departments of Immunology, University of Texas Southwestern Medical Center, Dallas, TX, 75235
2 Departments of Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX, 75235
3 Departments of Microbiology University of Texas Southwestern Medical Center, Dallas, TX, 75235
4 Microbiome Research Laboratory, University of Texas Southwestern Medical Center, Dallas, TX, 75235
5 Children's Health, Dallas, TX, 75235

Aim of our research is to develop NGS based diagnostic assays that work with small specimen volume and thus are suitable for young children. In the present study, we used 100ul-1ml blood specimen to profile potential microbes that can populate bloodstream of 1-2 year old children and trigger an immune response. We performed 16S rRNA amplicon sequencing on 200 pediatric blood specimens to detect bacterial DNA signatures. We also measured serum IgG antibody titers against a panel of pathogens, vaccine antigens, and common allergens. The blood DNA was extracted with ZymoBIOMICS kit followed by 16S V3-V4 sequencing on MiSeqDx and MinION sequencer. DADA2 plugin from QIIME2 was used for denoising, chimera removal and merging the forward and reverse reads. About 90% samples did not show any significant bacterial DNA signatures, consistent with the idea that blood is sterile environment. However, remaining 10% samples did show significant number (>20K) of 16S sequencing reads, suggesting a possible microbial signal. Taxonomic classification on this subset revealed that Firmicutes and Proteobacteria are predominant bacterial members communities in the pediatric blood. Staphylococcus, Deinococcus, Streptococcus, etc. were the most abundant genera observed. In addition, DNA signatures of some commensal microbes were also detected. Unsupervised clustering of sequencing data showed coaggregation of pathogens and commensals. 16S data showed strong signature of some commensals i.e. Veillonella and Prevotella DNAs in some samples. These bacteria are part of normal gut and oral flora in humans. Their presence in blood is interesting and may have a clinical relevance and infection prediction potential. Veillonella are Gram-negative, anaerobic cocci, known for their lactate fermentation abilities. Veillonella species are associated with oral infections and soft tissue & bone infections i.e. Osteomyelitis and Endocarditis. Similarly, Prevotella species are also member of normal gut and oral microbiome. This bacterium is also associated with respiratory infections and periodontal abscesses. Given that both of these commensal organisms have been reported to be part of oral microbiota and associated with bone and dental infections, it is possible that young children get these disseminated from oral cavity into bloodstream. We are performing whole genome sequencing on interesting specimens using long-read technology to generate genomescale data for more robust genomic characterization. Overlay of antibody data identified subjects with increased chances of transitioning to more severe clinical disease based on their heightened immune response. We hope to develop diagnostic assays that work with 1ml of blood and enable simultaneous characterization of blood infections and immune responses in 1-2 year old children.

Calypso: Web tools for collaborative, longitudinal genomic diagnostic care

POSTER 140

Alistair Ward[1,2], **Tony Di Sera**[1], **Stephanie Georges**[1], **Anders Pitman**[1], **Marti Tristani-Firouzi**[1], **Gabor Marth**[1,2]

1 University of Utah School of Medicine, Salt Lake City UT
2 Frameshift Labs, Inc., Cambridge MA

Patients presenting with complex phenotypes that evolve over time and requiring care from multiple medical disciplines represent some of the most challenging diagnostic cases. Achieving genetic diagnosis requires the combined effort of large diagnostic teams: a detailed understanding of phenotypes and family history is provided by the treating physician; the medical geneticist brings deep knowledge of genetic diseases; bioinformaticians provide computational analysis and adjudication of a variant's quality and pathogenicity; and diagnostic pathologists synthesize all information to draw conclusions on a variant's clinical significance. Existing software tools typically cater to those with computational expertise (e.g. bioinformaticians), rather than clinical expertise, and focus on performing diagnostic analysis at a single point in time. This results in two areas of concern: 1) treating physicians, or genetic counselors can be left out of the diagnostic process, limiting access to their unique insights, and; 2) a failure to follow a patient over time during a potentially lengthy diagnostic process. Many complex diagnostic cases are long-term processes that can continue for months or even years, and require regular attention to ensure changes in patient phenotypes and/or variant knowledge inform the diagnosis at the earliest opportunity.

We are developing Calypso, a comprehensive software system to address these concerns, and make team-based, longitudinal analysis a reality. Calypso comprises: 1) a computational backend with variant prioritization, re-annotation, and re-analysis pipelines; 2) iobio web tools to leverage each member's expertise, and summarize a patient's clinical and genomic history, and; 3) a platform to support close collaboration between all diagnostic team members. These intuitive apps are designed to encourage participation in the diagnostic process by the full team, and ensure the patient's evolving phenotypes, watchlists of candidate genes and variants, and relevant metadata can easily be shared with all collaborators. At-a-glance summaries of patient cases and histories are provided via dashboards, including views of patient case evolution. The University of Utah undiagnosed disease clinic and rapid NICU sequencing program serve as test programs for evaluation of Calypso, specifically chosen as rare disease clinics, and sequencing programs for critically ill newborns epitomize the team-based longitudinal diagnostic process.

Characterization of genome-wide regulatory networks in 150 iPSC lines provides novel insights into early embryonic self-renewal and development

POSTER 145

A D'Antonio-Chronowska[1], **TD Arthur**[1], **MF Pera**[2], **M D'Antonio**[1], **KA Frazer**[1]

1 UC San Diego, CA 92093
2 Jackson Laboratory, ME 04609

Co-expression networks have been widely used to identify genes with coordinated expression across biological samples. However, the regulatory mechanisms underlying transcriptome-wide coordinated gene expression are unknown. We hypothesize the existence of networks of genome-wide co-accessible regulatory elements, whose co-accessibility is mediated by differential transcription factor (TF) binding in subpopulations or cell states that vary in proportion across samples. Here, we integrate ATAC-seq in 150 induced pluripotent stem cell (iPSC) lines from 133 individuals with RNA-seq and WGS data in order to identify and characterize: 1) cell state-specific gene network modules (GNMs); and 2) genome-wide co-accessible regulatory element network modules (RNMs).

We initially examine transcriptome-wide co-expression across 16,110 genes and identify 68 GNMs (mean = 212 co-expressed genes; range 75 - 434). 25 GNMs are associated with 12 previously identified iPSC subpopulations (including enriched self-renewal (ESR) cells, 8 cell-like cells, and founder cells), which represent cell states that vary in proportion across iPSC lines.

We next examine genome-wide co-accessibility across 59,499 ATAC-seq peaks and identify 26 RNMs (mean = 1,892 co-accessible peaks; range 52 to 4,543), of which 22 are strongly associated with one or more GNMs and iPSC subpopulations i.e., the ESR subpopulation is associated with six RNMs harboring two distinct predicted TF binding sites (pTFBSs) profiles: 1) RNMs 4.1, 4.3, 4.4, and 4.5 are enriched for pluripotent and developmental TFs (NANOG, OCT4, TEAD1 and TEAD4) and enhancers; and 2) RNMs 2.2 and 2.3 are enriched with CTCF pTFBSs, bivalent promoters and repressed polycomb chromatin. These findings suggest that multiple cell state RNMs exist in the ESR subpopulation and that regulation of TAD formation and genes poised for expression play important roles in iPSC self-renewal. We perform a chromatin accessibility QTL (caQTL) and determine that peaks in certain RNMs are enriched for being affected by genetic variation while the peaks in other RNMs are depleted (ie RNMs are under different evolutionary constraint).

Our study is among the first to examine genome-wide regulatory co-accessible networks and provides novel insights into iPSC self-renewal and other important cell states. We also provide a framework for examining co-accessible regulatory networks in other cell types.

Accurate short variant calling from sequencing data with pangenomes and DeepVariant

POSTER 155

Mira Mastoras[3], **Mikhail Kolmogorov**[1], **Kimberley Billingsley**[2], **Melissa Meredith**[3], **Ryan Lorig-Roach**[3], **Pilar Alvarez Jerez**[2], **Laksh Malik**[2], **Kishwar Shafin**[4], **Trevor Pesout**[3], **Paolo Carnevali**[5], **Arang Rhie**[6], **Miten Jain**[7], **Sonja Scholz**[8], **Bryan Traynor**[9], **Winston Timp**[10], **Adam Phillippy**[6], **Fritz J Sedlazeck**[11], **Cornelis Blauwendraat**[2], **Benedict Paten**[3]

1 National Cancer Institute, National Institutes of Health, USA
2 Center for Alzheimer and Related Dementias, National Institutes of Health, USA
3 UC Santa Cruz Genomics Institute, Santa Cruz, CA, USA
4 Google LLC, 1600 Amphitheatre Pkwy, Mountain View, CA, USA
5 Chan Zuckerberg Initiative, Redwood City, CA, USA
6 National Human Genome Research Institute, National Institutes of Health, USA
7 Department of Bioengineering, Department of Physics, Northeastern University, Boston, MA
8 National Institute of Neurological Disorders and Stroke
9 National Institute of Aging, National Institutes of Health, USA
10 Biomedical Engineering, Johns Hopkins University, USA
11 Baylor College of Medicine, Houston, TX

The improved capability of long-read sequencing to detect genomic variation has enabled breakthrough advancements in human genetic research. Long-read sequencing was used to produce the first complete human genome assembly by the Telomere-to-Telomere consortium. Additionally, the Human Pangenome Reference Consortium recently released 30 nearly-complete haplotype-resolved human genomes from diverse backgrounds sequenced using multiple long-read technologies. Despite the promise long-read sequencing holds for uncovering complex variants, its cost remains a barrier for large-scale disease association studies. In this work, we developed a specialized DNA preparation protocol for Oxford Nanopore (ONT) PromethION to optimize the yield of a single flow cell. We combined it with a novel computational pipeline to produce haplotype-resolved de novo assemblies and characterize small and structural variants.

Using our DNA preparation protocol we produced ONT sequencing data for three cell lines (HG002, HG02723, HG00733) and 15 postmortem brain tissue samples from the National Institute of Aging collection. Each genome was sequenced using one flow cell, on average yielding 142 Gb of reads with read N50 30 kb. We performed reference-based small variant calls using PEPPER-Margin-DeepVariant, and found that within the GIAB confidence intervals our SNP calls outperformed Illumina on all three cell lines (HG002 F1-score 0.9976 vs 0.9965). In difficult-to-map regions, the improvement of ONT-based SNP calls over Illumina becomes more noticeable: HG002 MHC F1-score 0.9946 vs 0.9732, HG002 SegDup F1-score 0.9799 vs 0.9512. We used Shasta to assemble each genome into haploid contigs and developed a pipeline, Hapdup, that converts a haploid assembly into a pair of locally phased haplotypes. The resulting haplotype-resolved assemblies had mean NG50 ~20 Mb with base-level quality ~QV35. We produced structural variant calls via alignment to these assemblies, and they showed slightly better concordance with the GIAB callset (HG002 F1 score = 0.9555) compared to Sniffles2 (HG002 F1-score = 0.9523).

Our DNA protocol and computational workflow is cost effective and demonstrates competitive variant calling accuracy, opening the possibility for large scale long-read sequencing projects. It is currently being used to sequence 4000+ human brain samples at the Center of Alzheimer and Related Dementias at the NIH.

Conventional genomic predictive markers to immune checkpoint inhibitors are deteriorated in metastatic clear cell renal cell carcinoma

POSTER 160

Ahyeon Kim[1], **Eunyoung Kim**[1], **Minyong Kang**[2,3,4*], **and Sangwoo Kim**[1,*]

1 Department of Biomedical Systems Informatics, Graduate School of Medical Science, Brain Korea 21 Project, Yonsei University College of Medicine; 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, South Korea
2 Department of Urology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea;
3 Department of Health Sciences and Technology, SAIHST, Sungkyunkwan University, Seoul, Korea;
4 Samsung Genome Institute, Samsung Medical Center, Seoul, Korea
* co-correspondence

Immune checkpoint inhibitors (ICIs) have been one of the most effective cancer therapeutic options in first-line to advanced cancers, but demand effective biomarkers to complement the low responsiveness. Many genomic- and transcriptomic-level predictive markers to ICIs have been identified, including high tumor mutation (TMB) and copy-number variation (CNV) burdens, expression of immune checkpoint genes, and activated immunity measured in tumor microenvironment (TME), some to all of which are correlated with a better response in many different kinds of cancers, whereas no evidences were known in clear cell renal cell carcinoma (ccRCC). To test the applicability of conventional markers, we conducted DNA (targeted) - and RNA-sequencing of 60 metastatic ccRCC patients of two groups (26 responders; R and 34 non-responders; NR) in response to ICI therapy. There was no difference in the number of mutations (TMB) between the two groups (R vs. NR = 3.9 vs. 4.5 mut/Mb; p=0.42). Likewise, none of three immune checkpoint genes (PD-1, PD-L1 and CTLA-4) and 24 TME-related immune activity markers were differentially expressed. Rather, unexpected enrichment in the number of predicted neoantigen (on HLA-A; R vs. NR = 0.72 vs. 1.2; p=0.013), and number of genes affected by CNV (R vs. NR = 87 vs. 112; p=0.089) was shown in NR. We found that the only two features that discriminated the patient's response were VHL mutation status (R vs. NR = 85% vs. 56%; p=0.025) and MTOR amplification status (R vs. NR = 0% vs. 18%; p=0.031). Our results imply that immune environments and immunoediting mechanisms in ccRCC might be distinct from other cancers and the needs for development of novel markers to achieve accurate prediction of ICI response.

Development of functional brain organoids for disease modeling

POSTER 170

Alysson R. Muotri[1,2]

1 University of California San Diego, School of Medicine, Department of Pediatrics/Rady Children's Hospital San Diego, La Jolla, California 92093, USA.
2 University of California San Diego, School of Medicine, Department of Cellular & Molecular Medicine, La Jolla, California 92093, USA.

Structural and transcriptional changes during early brain maturation follow fixed developmental programs defined by genetics. However, whether this is true for functional network activity remains unknown, primarily due to experimental inaccessibility of the initial stages of the living human brain. We developed cortical organoids that spontaneously display periodic and regular oscillatory network events that are dependent on glutamatergic and GABAergic signaling. These nested oscillations exhibit cross-frequency coupling, proposed to coordinate neuronal computation and communication. As evidence of potential network maturation, oscillatory activity subsequently transitioned to more spatiotemporally irregular patterns, capturing features observed in preterm human electroencephalography (EEG). These results show that the development of structured network activity in the human neocortex may follow stable genetic programming, even in the absence of external or subcortical inputs. Our approach provides novel opportunities for investigating and manipulating the role of network activity in the developing human cortex. Applications for neurodevelopmental disorders will be discussed.

EagleC: Detecting structural variants in cancer genomes from bulk or single-cell Hi-C

POSTER 175

Xiaotao Wang, Yu Luan, Feng Yue

Northwestern University

Hi-C technique has been 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 intra- chromosomal SVs (>1Mb) 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. Importantly, we show that EagleC can uniquely capture a set of fusion genes that are missed by WGS or nanopore. Furthermore, EagleC also effectively captures SVs in other chromatin interaction platforms, such as HiChIP, ChIA-PET, and capture Hi-C. We apply EagleC in over 100 cancer cell lines and primary tumors, and identify a valuable set of high-quality SVs. Finally, we demonstrate that EagleC can be applied to single-cell Hi-C and used to study the SV heterogeneity in primary tumors.

Enabling all four facets of P4 medicine from the standpoint of the pathology tissue specimen

POSTER 180

Wilfrido Mojica[1], **Yun Wu**[2], **Troy Wood**[3], **Natesh Parashurama**[4], **Supriya Mahajan**[5]

1 Department of Pathology and Anatomical Sciences, University at Buffalo, Buffalo, NY
2 Department of Biomedical Engineering, University at Buffalo, Buffalo, NY
3 Department of Chemistry, University at Buffalo, Buffalo, NY
4 Department of Chemical and Biological Engineering, University at Buffalo, Buffalo, NY
5 Department of Medicine, University at Buffalo, Buffalo, NY

The vision that the field of medicine will transition from reactive disease care to a more predictive, preventative, personalized and participatory (P4) model was conceived over a decade ago. This model is thought to have the potential to drive profoundly beneficial health-care related economics, policies and social change. While there has been progress towards this goal, one key, fundamental core component has not changed, and that is how tissue biopsy specimens are processed. The current approach immerses tissue in formalin, is decades old, molecularly unfriendly, and results in a static, non-viable specimen largely restricted to brightfield visual examination. Yet it is the current state-of-the-art in pathology laboratories and so cannot be discontinued because of its integral role in diagnosis and tumor staging. Although this purpose is important, we are of the belief that the potential informative data that can be mined from a tissue biospecimen is extensive and untapped. To that end, we have developed a simple, labor un-intensive, nominally resource needy, pre-analytic modified processing approach to recover analytes and cells for a myriad of downstream testing purposes. An important caveat is that the originating tissue is left intact and so remains available for the needs of clinical diagnostics. This approach improves upon the current passive means of processing tissue biospecimens, yielding exosomes, small metabolites and dislodged cells from the parent tissue that can be used for biomarker identification, tumor subtyping, prognostication and the development of organoids for personalized therapeutics. Herein we intend to demonstrate that the recovery of these substrates, in combination with improvements in sample integrity and accessibility, and through the use of standard and novel laboratory tools, have the potential to obviate the persistent roadblock associated with tissue biospecimens and the integration of P4 medicine, as a whole, into clinical practice.

Fine mapping spatiotemporal mechanisms of genetic variants underlying cardiac traits and disease

POSTER 185

Nayara Sousa[1], **Jennifer Phuong Nguyen**[1], **Agnieszka D'Antonio-Chronowska**[1], **Kelly A. Frazer**[1], **Matteo D'Antonio**[1]

1 UC San Diego, CA 92093

Although Genome-wide association studies (GWAS) have identified thousands of loci associated with cardiac traits and diseases, for most of these associations the underlying causal variants have not yet been delineated. All the approaches utilized to date to identify causal variants and their associated molecular mechanisms underlying cardiac GWAS loci have strong limitations and no large-scale fine mapping efforts have been published. Identifying putative causal variants and understanding their associated molecular mechanisms, such as their target genes, is further complicated by the fact that gene expression is regulated in a spatiotemporal-specific manner. In this study, we performed an eQTL analysis on 966 cardiac RNA-seq samples from multiple adult cardiac-related tissues from GTEx (atrial appendage, left ventricle, aorta and coronary artery) and fetal-like samples from our previously published iPSCORE collection (iPSC-derived cardiovascular progenitor cells, iPSC-CVPC), and detected eQTL signals associated with more than 11,000 eGenes and 7,000 eIsoforms. By comparing the eQTL signals between samples at each stage (fetal-like or adult) or annotated as a specific organ (arteria and heart) or tissue (atrium, ventricle, aorta and coronary artery) against all the other samples and by deconvoluting the cell types in each of the 966 cardiac samples, we were able to annotate 2,578 eQTL signals as associated with stage, organ, tissue and/or cell type. To identify the potentially causal variants for cardiac traits and diseases, we obtained summary statistics for five cardiac GWAS from the UK BioBank (pulse rate, pulse pressure, QRS interval, atrial fibrillation and myocardial infarction) and colocalized their signals with eQTLs. We observed that three of the five cardiac traits are enriched for being functional in specific spatiotemporal contexts, including associations between pulse pressure and fetal-like-eQTLs, pulse rate and adult-eQTLs, and atrial fibrillation and cardiac muscle. We also used the colocalizations with eQTLs to fine map 210 cardiac GWAS loci and determined that, of 79 fine-mapped loci (credible sets with ⩽ 5 SNPs), 15 had spatiotemporal eQTL signals.

In summary, we show that many genetic variants likely affect the expression of their associated genes and, in turn, complex cardiac traits and disease in a stage-, tissue- and cell type-specific fashion.

Framework for therapeutic target discovery and multi-omic–supported individualized medicine

POSTER 190

Cynthia Lawley[1], **Philippa Pettingill**[2], **Klev Diamonti**[2], **Lotta Wik**[2], **Niklas Nordberg**[2], **John Broberg**[2], **Johan Björkesten**[2], **Erika Assarsson**[2], **Sara Henriksson**[2], **Ida Grundberg**[2], **Christina Westerberg**[2], **Elin Liljeroth**[2], **Adam Falck**[2], **Martin Lundberg**[2], **Lasse Folkersen**[3], **Anders Mälarstig**[4]

1 Olink, Population Health, San Francisco, USA
2 Olink, Uppsala, Sweden
3 Nucleus Genomics, Ltd New York, USA
4 Karolinska Institutet, Stockholm, Sweden

Understanding the dynamics of the human proteome is crucial for identifying biomarkers to be used as measurable indicators for disease severity and progression, patient stratification, and drug development. These data also amplify power to detect genetic support therapeutic target discovery. The Proximity Extension Assay (PEA) is a technology that translates protein information into actionable insights across large samples sizes in both healthy and disease samples. The high-throughput nature of the assay is enabled by linking protein-specific antibodies to DNA-encoded tags that can be read out on a next generation sequencer. We have combined the PEA technology described above with automated sample preparation and a high-throughput sequencing readout for parallel measurement of ~3,000 proteins for up to 384 samples at a time, generating over 1 million data points per run. Characterizing the proteome alongside genetic and clinical data enables a pQTL framework to not only validate known clinical targets and identify new clinical targets but to also suggest repurposing opportunities of clinical candidates for new indications. Here we will summarize results where proteomics is impacting large population health studies (e.g., UK Biobank, SCALLOP) to advance precision and personalized medicine.

Genetics providers are key to the integration of genetic testing within the practice of frontline clinicians

POSTER 200

Maren T Scheuner[1,2], **Paloma Sales**[1], **Katherine Hoggatt**[1,2], **Ning Zhang**[1], **Mary A Whooley**[1,2], **Michael J Kelley**[3]

1 San Francisco VA Health Care System
2 UCSF School of Medicine
3 VA National Oncology Program Office

Objective: The number of genetics professionals has not kept pace with the demand for genetic services, and near-term, there is limited ability to grow the genetics workforce. Frontline clinicians are key to ensuring access to genetic testing. We sought to understand use of genetic tests by frontline clinicians within the Department of Veterans Affairs (VA).

Methods: We administered a web-based survey to clinicians at 20 VA facilities. We assessed clinician genetic testing experience, preferences for genetic testing roles/responsibilities, and genetics education needs. Physicians, nurse practitioners (NPs), physician assistants (PAs), and pharmacists were eligible. We excluded genetics professionals and clinicians not seeing patients. Multiple logistic regression assessed associations between clinician characteristics and currently ordering genetic tests.

Results: The response rate was 11.3% (1,207/10,680) with 909 eligible. Most (63.0%) were physicians, 21.9% NPs, 10.1% pharmacists, and 5% PAs. Only 20.8% reported feeling prepared to use genetic tests and 13.0% were currently ordering genetic tests, though usually only one or two a year. Delivery of genetic tests without involving genetics was preferred by only 7.9%. Characteristics positively associated with ordering genetic tests included believing improving genetics knowledge could alter their practice (OR=1.87,95%CI:1.29-2.66,p=.001), feeling prepared to use genetic tests (OR=6.91,95%CI:4.33-11.03,p<.001), and referral of at least one patient to genetics in the past year (OR=3.23,95% CI:1.53-6.84,p.002). Pharmacists were less like to order genetic tests vs. primary care (OR=0.16,95%CI:0.07-0.37,<.001). On average, 17.2% had genetics education in the past year. Top-rated educational priorities included common genetic disorders relevant to their practice, recognizing patients who may benefit from genetic testing, pharmacogenetic tests to inform treatment, and knowing different types of genetic tests. The clinical genetics team was the preferred choice for providing updates on genetics.

Conclusions: Genetic testing is infrequently used by frontline clinicians in the VA and these clinicians remain underprepared to use genetic tests in their practice. As clinicians use genetic testing in their practice, there is a paradoxical positive association with genetics referrals.

Impact: The demand for genetics providers should increase as frontline clinicians integrate genetics into their practice. These findings should be considered in workforce planning, and the education and training of frontline clinicians.

Genome reference impacts RNA-seq interpretation and rare disease diagnosis

POSTER 205

P. Goddard[1], **R. Ungar**[1], **T. Jensen**[1], **K. S. Smith**[1], **C. Jin**[1], **D. Bonner**[2], **S. Marwaha**[1], **J.A. Bernstein**[3], **M. Wheeler**[1], **S. B. Montgomery**[1]

1 Stanford Univ., Stanford, CA
2 Stanford Ctr. for Undiagnosed Diseases, Stanford, CA
3 Stanford Univ. Sch. of Med., Stanford, CA

Transcriptome analysis is a powerful tool for unraveling molecular mechanisms and improving Mendelian disease diagnostics; however, despite evidence that reference genome selection directly impacts variant identification and diagnostic yield from exome and genome sequencing data, no study to date has systematically assessed the impact of reference genome selection on RNA-seq analysis. We present the impact of the hg19, hg38, and CHM13 reference genomes on expression, splicing, and outlier detection in a rare disease cohort.
We generated paired-end RNA-seq data on blood samples from 330 rare disease patients and family members from the Undiagnosed Diseases Network (UDN), with primarily neurological, musculoskeletal, or immune-related conditions. Data were aligned to the primary genome assemblies, using the GENCODEv35 equivalent gene annotations for hg38.p13, hg19.p13, and CHM13v2. To identify genes with substantial sequence changes between builds, sequence similarity was calculated from pairwise alignment of all exons per gene. We subsequently quantified gene, transcript and splice junction expression, and identified patient-specific expression and splicing outliers.

Pairwise alignment revealed that 99% of genes annotated in both hg19 and hg38 had above 95% sequence similarity across exons, compared to just 12% of genes annotated in hg38 and CHM13. The top 5% of genes with the greatest discrepancies in expression between hg19 and hg38 include 46 genes previously shown to be enriched for variants discrepant between the builds by Li et al (PMID 34129815), and 93 known OMIM genes, including INF2, which underlies a form of Charcot-Marie-Tooth disease. We then investigated how build-induced differences in quantification impacted our interpretations for diagnosed rare disease cases. Seven out of 343 Mendelian disease genes implicated in solved cases across UDN sites had significantly different expression between hg19 and hg38 in our cohort, demonstrating that build can impact RNA-seq guided diagnosis. We further found that overall only 52% of expression outliers and 73% of splicing outliers were consistent between the two builds. We will present similar analyses investigating the impact of switching from hg38 to CHM13.

RNA-seq is a widespread method, yet the understanding of how reference genome selection impacts both quantification results and Mendelian disease diagnosis has not been investigated. We present RNA-seq data on a large number of rare disease patients and demonstrate that in the context of a maximally consistent gene annotation, reference genome selection impacts quantification and clinical interpretation of transcriptomic data.

Genomic and transcriptomic profiling reveals molecular characteristics of parathyroid carcinoma

POSTER 210

Se-Young Jo[1,†], **Namki Hong**[2,†], **Seunghyun Lee**[2,3], **Jong Ju Jeong**[4], **Jeongsoo Won**[1], **Jiho Park**[1], **Gi Jeong Kim**[5], **Sang Kyum Kim**[5], **Sangwoo Kim**[1,*], **Yumie Rhee**[2,*]

1 Department of Biomedical Systems Informatics and Brain Korea 21 PLUS Project for Medical Science, Yonsei University College of Medicine, Seoul, Korea
2 Department of Internal Medicine, Severance Hospital, Endocrine Research Institute, Yonsei University College of Medicine, Seoul, Korea
3 Department of Internal Medicine, Wonju Severance Christian Hospital, Yonsei University Wonju College of Medicine, South Korea
4 Department of Surgery, Severance Hospital, Yonsei University College of Medicine, Seoul, Korea
5 Department of Pathology, Yonsei University College of Medicine, Seoul, Korea

†These authors contributed equally to this work.
* Corresponding authors

Genomic and transcriptomic profiling has enhanced the diagnostic and treatment options for many cancers. However, the molecular characteristics of parathyroid cancer remain largely unexplored, thereby limiting the development of new therapeutic interventions. Herein, we conducted genomic and transcriptomic sequencing of 50 parathyroid tissues (12 carcinomas, 28 adenomas, and 10 normal) to investigate the intrinsic and comparative molecular features of parathyroid carcinoma. We confirmed multiple two-hit mutation patterns in Cell Division Cycle 73 (CDC73) that converged to bi-allelic inactivation, calling into question the presence of a second hit in other genes. In addition, allele-specific repression of CDC73 in copies with germline truncating variants suggested selective pressure prior to tumorigenesis. Transcriptomic analysis identified upregulation of the expression of E2F targets, KRAS and TNF-alpha signaling, and epithelial-mesenchymal transition pathways in carcinoma compared to that in adenoma and normal tissues. A molecular classification model based on carcinoma-specific genes clearly separated carcinoma from adenoma and normal tissues, the clinical utility of which was demonstrated in two patients with uncertain malignant potential.
A deeper analysis of gene expression and functional prediction suggested that Wilms Tumor 1 (WT1) is a potential biomarker for CDC73-mutant parathyroid carcinoma, which was further validated through immunohistochemistry. Overall, our study revealed the genomic and transcriptomic profiles of parathyroid carcinoma and may help direct future precision diagnostic and therapeutic improvements.

Genomic and transcriptomic profiling reveals molecular characteristics of parathyroid carcinoma

POSTER 215

Isaac Joshua[1], **Arya Massarat**[1], **Jonathan Margoliash**[1], **Melissa Gymrek**[1], **Kelly A Frazer**[1], **Matteo D'Antonio**[1]

1 UC San Diego, CA 92093

The 4 Mb major histocompatibility (MHC) region on chromosome 6p21.3, which encodes the human leukocyte antigen (HLA) gene complex, is highly polymorphic, gene dense (383 genes), and has strong linkage disequilibrium (LD). GWAS have identified many traits and diseases with significant associations; however, ~90% of GWAS have been performed on Europeans and have primarily focused on common SNPs. HLA genes may have thousands of alleles, which are categorized as HLA types and are poorly captured by GWAS utilizing SNPs; therefore, there are likely many genetic associations in the MHC region not yet discovered.

We utilized 479,806 individuals in the UK BioBank to examine associations between the genotypes of 346 imputed HLA types at 4-digit resolution (which distinguish nonsynonymous variants between alleles) for 11 HLA genes and >75,000 SNPs in the MHC region and 1,213 traits, including: diseases; blood biochemistry markers; and hematological parameters on whole blood. We observed genome-wide significant associations for 419 traits. Although >90% of UKBB individuals were from European ancestry, only 60 traits were European-specific, whereas 44 were specific for individuals from African descent and 70 for East Asians, indicating that the MHC region has strong ancestry-specific components.

To understand the differences between genetic associations for 4-digit HLA types and SNPs, we performed fine-mapping using susieR on each genome-wide significant trait and found 2,215 independent signals. For 37 signals, the variant with the strongest posterior inclusion probability was an HLA type, indicating that HLA types provide information beyond SNP genotypes. These signals include the previously reported association between HLA-DRB1*01:03 and ulcerative colitis across all individuals, as well as ancestry-specific associations, including: HLA-C*15:02 and kidney failure in Africans; HLA-DRB1*04:04 and aortic valve disorders in East Asians; HLA-B*35:01 and lymphocyte percentage in Europeans; and HLA-B*14:01 and C-reactive protein levels in South Asians.

Our study shows that the MHC region is associated with hundreds of traits in an ancestry-specific fashion, and that many of these associations are likely driven by regulatory variants. Expanding HLA types from 4-digit to 8-digit resolution (which distinguish nonsynonymous, synonymous, and non-coding variants between alleles) will likely identify hundreds more associations with specific HLA gene alleles.

Improving cardiovascular diagnostic yield with long-read technology

POSTER 225

Sara E Kalla[1], **Medhat Mahmoud**[1], **Qingchang Meng**[1], **Ginger A Metcalf**[1], **Donna M Muzny**[1], **Fritz J Sedlazeck**[1], **Christie M Ballantyne**[2], **Richard A Gibbs**[1]

1 Baylor College of Medicine, Human Genome Sequencing Center
2 Baylor College of Medicine, Molecular and Human Genetics

The promise of genomic medicine is to deliver an accurate prediction of how one's genome will impact health via reliably identifying all contributing variants. We are particularly focused on how long-read technology improves calling and diagnostic yield for more complex variants such as LDLR and Lp(a). Twenty-one individuals who previously were recruited for a cardiovascular panel study (HeartCare, Murdock et al 2021 https://doi.org/10.1038/s41436-021-01294-8) underwent whole genome sequencing (WGS) via long-read Oxford Nanopore Technology (ONT). LDLR is well known for having Copy Number Variants (CNVs) that affect gene expression, however, correctly identifying the structure of LDLR is complicated by numerous Alu repeat elements that contribute to structural diversity. Long-read technology is ideal for mapping to such regions; using long reads, all twenty-one individuals are phased correctly.

While the HeartCare panel also included the gene LPA, technical limitations with short-read sequencing limited the reporting to the known risk SNV, rs41272114. We are in the process of evaluating coverage, phasing and copy number across the 5.5 kbp Kringle repeats (KIV) in LPA. Lower repeat number of KIVs has been correlated with higher Lp(a) plasma concentrations and thus increased risk of myocardial infarction. Correctly identifying the repeat number will add to the utility of cardiovascular genomic testing. In my presentation, I will highlight some of the ways that long-read technology will contribute to both clinical research and genomic medicine.

Large-scale mutagenesis of mammalian enhancers in vivo

POSTER 230

Michael Kosicki[1], **Stella Tran**[1], **Veena Afzal**[1], **Jennifer A. Akiyama**[1], **Ingrid Plajzer-Frick**[1], **Catherine S. Novak1**, **Yiwen Zhu**[1], **Momoe Kato**[1], **Anne Harrington**[1], **Riana D. Hunter**[1], **Kianna von Maydell**[1], **Janeth Godoy**[1], **Eman M. Meky**[1], **Sarah Barton1**, **Erik Beckman**[1], **Diane E. Dickel**[1], **Axel Visel**[1], **Len A. Pennacchio**[1]

1 Lawrence Berkeley National Lab, Berkeley, CA

Mutations in regulatory DNA account for a large fraction of variants suspected to contribure to human diseases. However, interpretation of these variants is hampered by the lack of accurate models of regulatory DNA activity. Systematic mutagenesis of enhancers in vivo is necessary to understand their logic. Such studies have so far been hampered by the lack of high throughput tools to interrogate sequence perturbations and the prohibitively large size of most in vivo validated regulatory elements.

To counter these challenges, we have isolated a set of seven small enhancers (200-400 bp) that reliably promote tissue-specific expression in developing mouse embryos, and applied our novel transgenic assay enSERT strategy to test at scale the in vivo consequences of mutagenizing 167 blocks of 12 bp. Architecture of individual enhancers was revealed to be idiosyncratic, with some elements being very robust to mutations (no loss-of-function blocks, many blocks can be mutagenized without consequences) and others carrying regulatory information in most blocks tested. We also combined mutations to reveal epistatic interactions, mapped out mutations that cause separation-of-function (selective loss of a structure) and found repressor elements that lead to gain-of-function when mutagenized. Our results underscore the complexity of regulatory elements and provide a foundation for validation of computational approaches to regulatory variant prediction.

Mutations in Histone H3 are associated with immunosuppressive signatures in pediatric high-grade gliomas

POSTER 240

Wesley Kwong[1], **Mike Schachter**[1], **Samantha Liang**[1], **Sandra Santulli-Marotto**[1], **Lacey Padron**[1], **Robert Prins**[3], **Derek Oldridge**[2,] **and Dinesh Kumar**[1]

1 Parker Institute for Cancer Immunotherapy
2 Children's Hospital of Philadelphia, Department of Pathology and Laboratory Medicine
3 David Geffen School of Medicine at UCLA, Departments of Neurosurgery and Pharmacology

Gliomas are common, histologically heterogeneous pediatric brain tumors. Five-year survival ranges from <10% for diffuse intrinsic pontine gliomas to >90% for focal pilocytic astrocytomas. Immunotherapy efficacy has been impeded by tumor heterogeneity, acquired and adaptive resistance, and the brain's immunosuppressive tumor microenvironment.
Histone H3 is a common epigenetic mutation in pediatric gliomas. We investigated differences in somatic variants for patients with a deleterious H3 mutation (n=9) vs H3 wildtype (n=16), relating these differences to their bulk transcriptomes.

Differential expression analysis showed lower expression for LHX2 in H3 mutants (log2FC=-1.58, p=0.002). LHX2 regulates the SLIT/ROBO pathway, which influences macrophage chemotaxis. We noted higher expression of SLIT2 in H3 mutants (log2FC=1.12, p=0.001). SLIT2 inhibits macrophage invasion, and immunotherapy is more efficacious in SLIT2 knockout mice.

Pathway analysis of somatic mutations showed that H3 mutants had a higher likelihood of a mutation in the IFNG pathway, even after adjusting for TP53 mutation status (logodds=7.2, 95% confint=[1.0, 69.2], p=0.057, n=25).

Gene expression analysis via signature scoring showed macrophage, IFNG response, and Verhaak mesenchymal glioma signatures correlated with Pearson coefficients >0.7. IFNG pathway mutations negatively correlated with these scores, ranging from -0.25 to -0.3. Mutations in the INFG pathway associated with decreased IFNG pathway activation (regression, p=0.075, n=25).

In summary, we showed that pediatric patients with H3 mutations exhibited reduced expression of LHX2 and increased expression of SLIT2, which may relate to reduced macrophage chemotaxis. They are also more likely to have mutations in the IFNG response pathway and reduced gene expression in that pathway. These effects may be implicated in immune suppression and immunotherapy resistance.

New delivery models improve access to germline testing for patients with advanced prostate cancer

POSTER 245

Maren T Scheuner[1,2], **Paloma Sales**[1], **Katherine Hoggatt**[1,2], **Samuel Washington**[1,2], **Eva Ferino**[1], **Christine N Serway**[3], **Sara Ahmed**[3], **Michael J Kelley**[3]

1 San Francisco VA Health Care System
2 UCSF School of Medicine
3 VA National Oncology Program Office

Objectives: The Department of Veterans Affairs (VA) launched a clinical pathway including both tumor and germline testing to inform treatment for advanced prostate cancer on 5/3/2021. Anticipating increased germline testing demand, new delivery models were created to augment the existing traditional model of referring patients to genetics providers. This included a non-traditional model where oncologists perform all pre- and post-test activities and consult genetics when needed, and a hybrid model where oncologists obtain informed consent and place genetic consults for germline test ordering, results disclosure, and genetics follow-up, as needed. We sought to assess germline testing by delivery model.

Methods: Data sources included the VA National Precision Oncology Program dashboard and reports from contracted germline testing laboratories. Patient inclusion criteria: living as of 5/2/2021 with VA oncology visits after 5/2/2021. Multivariate regression assessed associations between patient characteristics and germline testing between 5/3/2021 (pathway launch) and 5/2/2022, accounting for clustering of patients within ordering clinicians.

Results: We identified 16,041 patients from 129 VA facilities with average age 75 (SD=8.2,36-102), 29% Black and 60.% White. Throughout the first year, 896 (6%) patients had germline testing ordered by 60 clinicians at 67 facilities with 52% of orders by the hybrid model, 32% the non-traditional model, and 16% the traditional model. Patient characteristics positively associated with germline testing included care at hybrid model (OR=6.03,95%CI:4.62-7.88) or non-traditional model facilities (OR=5.66,95%CI:4.24-7.56) compared with the traditional model, completing tumor molecular testing (OR=5.80,95%CI:4.98-6.75), and Black compared with White race (OR=1.24,95%CI:1.06-1.45). Compared to patients aged <66, patients aged 66-75 and 76-85 were less likely to have germline testing (OR=0.74,95%CI:0.60-0.90 and OR=0.67,95%CI:0.53-0.84, respectively).

Conclusions: Though only 6% of advanced prostate cancer patients had germline testing since the pathway launch, the new delivery models were instrumental to improving access. Evaluation is ongoing to understand observed demographic differences in germline testing and the effectiveness of implementation strategies to promote adoption of the new delivery models.

Impacts: Continued spread and uptake of the hybrid and non-traditional models for germline testing should result in improve outcomes for advanced prostate cancer patients by increasing use of treatments targeting molecular findings and informing clinical trials eligibility.

Phenome-wide Mendelian randomization study of plasma triglycerides and 2,600 disease traits

POSTER 255

Joshua K. Park[1,2,3,*], **Shantanu Bafna**[1,2,*], **Iain S. Forrest**[1,2,3], **Áine Duffy**[1,2], **Carla Marquez-Luna**[1,2], **Ben O. Petrazzini**[1,2], **Ha My Vy**[1,2], **Daniel M. Jordan**[1,2], **Marie Verbanck**[4], **Jagat Narula**[5,6], **Robert S. Rosenson**[5,7], **Ghislain Rocheleau**[1,2], **Ron Do**[1,2,†]

1 Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA
2 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
3 Medical Scientist Training Program, Icahn School of Medicine at Mount Sinai, New York, NY, USA
4 Université Paris Cité, UR 7537 BioSTM, Paris, France
5 Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY
6 Cardiovascular Imaging Program, Zena and Michael A. Wiener Cardiovascular Institute, Mount Sinai Heart, Icahn School of Medicine at Mount Sinai, New York, NY
7 Metabolism and Lipids Unit, Zena and Michael A. Wiener Cardiovascular Institute, Mount Sinai Heart, Icahn School of Medicine at Mount Sinai, New York, NY
* These authors contributed equally to this work.
† Corresponding author

Triglycerides (TG) play vital roles in physiology, ranging from energy storage and mobilization to inflammation and thrombosis. However, controversy mires whether the relationship between plasma TG levels and disease risk is causal or associative in most human diseases. This topic remains contentious even in atherosclerotic cardiovascular disease (ASCVD), as reflected by the conflicting results of two recent landmark trials, STRENGTH and REDUCE-IT. Establishing causality between TGs and phenome-wide disease risk could inform drug development of the novel repurposing opportunities and adverse effects of approved and emerging TG-lowering therapies, including icosapent ethyl and APOC3 inhibitors. Clinical trials are currently evaluating TG-lowering therapies targeting ANGPTL3, ANGPTL4 and APOC3 for dyslipidemias.

Here, we conducted a phenome-wide, two-sample Mendelian randomization (MR) analysis using an inverse-variance weighted (IVW) estimator to infer the causal effects of plasma TG levels on 2,600 disease traits in the European ancestry population of UK Biobank. We then externally tested 221 nominally significant associations ($p < 0.05$) for replication in an independent population from FinnGen. To account for potential horizontal pleiotropy, we performed sensitivity analyses using MR-Egger, weighted median, and MR-PRESSO methods. Finally, we applied multivariable MR to isolate the independent effects of plasma TG levels from correlated lipid fractions.

Our results identified 7 disease traits reaching Bonferroni-corrected significance in both the discovery ($p < 1.92 \times 10\text{-}5$) and replication analyses ($p < 2.26 \times 10\text{-}4$), supporting a causal relationship between plasma TG levels and ASCVDs, including coronary artery disease (OR 1.33, 95% CI 1.24-1.43, $p = 2.47 \times 10\text{-}13$). We also identified 12 disease traits that were Bonferroni-significant in the discovery or replication analysis and at least nominally significant in the other ($p < 0.05$), identifying plasma TG levels as a novel risk factor for 9 non-ASCVD disease traits, including uterine leiomyoma (OR 1.19, 95% CI 1.10-1.29, $p=1.17 \times 10\text{-}5$).

Using a phenome-wide Mendelian randomization approach, we identified 19 disease traits with positive causal associations with plasma TG levels, many of which are outside the context of cardiovascular disease. This suggests the adverse side effects of approved or emerging TG-lowering agents or provides rationale for repurposing them towards novel indications.

Quantifying regional DNA methylation improves the detection of biologically relevant associations in Alzheimer's disease

POSTER 260

Tiffany Eulalio, **Daniel Nachun, Stephen B. Montgomery**

Stanford University

DNA methylation is an epigenetic modification linked to numerous neurological diseases, such as Alzheimer's Disease (AD), however, a mechanistic understanding of methylation's role in these diseases remains elusive. DNA methylation at individual CpG sites is often aggregated across functional genomic regions, such as promoters or enhancers, to improve interpretability. Oftentimes, the average methylation value is used for downstream analyses such as detecting differential methylation or unsupervised learning. While this can provide valuable insights, averaging regional methylation to a single value is noisy and further fails to capture more complex methylation patterns, particularly in regions with high CpG density or a mixture of hypo- and hypermethylated CpGs. We hypothesized that these patterns could be better captured by using principal components analysis (PCA) to summarize regional methylation while retaining orthogonal and interpretable methylation signals for downstream analyses. We demonstrated the utility of this approach in the Religious Order Study/Memory and Aging Project (ROSMAP) dataset, a large dataset with methylation collected from the dorsolateral prefrontal cortex (DLPFC) of AD patients and controls. In addition to analyzing the bulk tissue data, we used cell type deconvolution to estimate the cell type-specific methylation signal for four major brain cell types; astrocytes, endothelial cells, neurons, and oligodendrocytes/oligodendrocyte progenitor cells. We compared the performance of the regional aggregation of CpGs with principal components to averaging in identifying differentially methylated genes (DMGs) and pathway enrichment of DMGs for AD-relevant biology. Regional PCA identified over ten times more differentially methylated genes than averaging in the bulk data and up to four times more in the cell type-specific datasets. The genes identified by regional PCA included over 88 AD-relevant genes found in the Open Targets database (overall association score > 0.08, 90th percentile) that were not identified using averages; including the highly-evidenced AD genes APP, ABCA7, and TREM2. In conclusion, we show that summarizing DNA methylation within genomic regions using PCA captured more biologically relevant signals than averaging and improves our understanding of the role of methylation in disease.

Secondary findings as a model system for genomic screening

POSTER 265

Julie C. Sapp, Katie L. Lewis, Emily Modlin, Alana Davidson, Alexis Darling, Jennifer Johnston, Leslie G. Biesecker

Center for Precision Health Research
National Human Genome Research Institute
National Institutes of Health

Opportunistic screening for medically-actionable secondary genomic findings (SF) shares critical attributes with population screening, yet few data exist to support the effectiveness of either approach. We implemented a program to investigate the effectiveness of SF return, refine the SF definition, and evaluate understanding and downstream health care implementation with an eye toward population screening challenges and opportunities. Enrollment in the study began in January, 2020 and we received >700 inquiries from clinicians, biobanks, genetic testing companies, support groups, research studies, and self-referrals. In 1/3 of these, the finding was diagnostic/primary and not secondary. As in other settings, screening cannot replace diagnostic testing and understanding these misunderstandings of SF is critical for proper care. We present initial findings on these cases. We refined the definition of SF for the purpose of research aimed at understanding phenotypic prevalence, adherence to recommendations, and other outcomes of SF receipt. After eligibility screening, we have formally enrolled 128 recipients of genomic results that meet our refined definition of SF. These 128 cases represent 29 genes and 15 disorders. Preliminary data show that most had good understanding of evaluation recommendations for the SF. Only a handful reported engaging in a recommended health behavior two or more times since SF receipt. Participants reported disclosing their SF to first-degree relatives (81%) and 27% of these underwent cascade testing. These data show that there is a critical need for emphasizing the distinction of SF from primary findings in clinical practice. We encountered numerous cases where genomic testing was clearly indicated but not performed; opportunistic screening is an inappropriate testing approach for these individuals. The data also suggest that implementing effective and ongoing health care for individuals with SF requires greater clinician awareness and directiveness, more effective communication, and readily available care pathways to maximize the clinical utility of SF, which will likely be similarly needed in population screening.

Selection acting on somatic structural variation in blood impacts molecular function and cancer risk among humans

POSTER 270

Kimberly Skead[1,2], **David Soave**[3], **Vanessa Bruat**[1], **Michelle Harwood**[1,2], **Quaid Morris**[4], **Marie-Julie Fave**[1], **and Philip Awadalla**[1,2,5]

1 Ontario Institute for Cancer Research; Toronto, Canada.
2 Department of Molecular Genetics, University of Toronto; Toronto, Canada
3 Department of Mathematics, Wilfrid Laurier University; Guelph, Canada
4 Computational and Systems Biology Program, Memorial Sloan Kettering Cancer Center, New York, United States
5 Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

The age-associated accumulation of somatic point mutations in blood, termed clonal hematopoiesis of indeterminate potential (CHIP), has been implicated in the development of blood cancers and cardiovascular conditions. Here, we evaluate how natural selection shapes the prevalence of clonal somatic structural variants (SSVs) in blood, and their impact on cancer risk, among 13,760 individuals from the Canadian Partnership for Tomorrow's Health. We find that SSV-inferred CHIP is three times higher than previously reported with one in eight individuals harboring a SSV. Using a novel statistical approach, we identify SSV hotspots across the genome, and find that hotspots are enriched for genes implicated in blood cancer development. Correspondingly, individuals who harbor at least one SSV are 1.68x more likely to progress to blood cancer. The risk of progressing to cancer scales with the number of events per individual with 1-2 and 3+ SSVs conferring a 1.77x and 5.86x increase in risk of progressing to blood cancer, respectively. Interestingly, the predictive potential of SSVs is not limited to blood cancer as we find that individuals who harbor at least one SSV in their blood are 1.57x more likely to progress to breast, prostate, and pancreatic cancer, suggesting that the accumulation of blood-based SSVs might reflect critical immunological changes arising during early tumorigenesis. To determine the impact of SSVs on blood-based molecular phenotypes, we investigate the relationship between SSVs and the transcriptome. We show that gains, losses, and copy-number-neutral variants impact gene expression distinctly, with stabilizing selection shaping the penetrance of SSVs among blood transcriptomes. SSV-adjusted eQTL mapping finds that one in three canonical eQTLs are confounded by SSVs, revealing a previously overlooked somatic contribution to blood expression profiles, regulatory mechanisms, and disease development. Similarly, we show that SSV-loss events account for 9.1% of observed allele specific expression in whole blood transcriptome profiles. In summary, our work finds that not all SSVs impact molecular phenotypes equally, and that consideration of the multiple selective pressures acting on SSVs at the DNA and RNA level is critical when leveraging the predictive power of somatic events for early cancer detection.

SHINE: protein language model-based prediction of pathogenicity for InFrame insertion and deletion variants

POSTER 275

Xiao Fan[1,2], **Hongbing Pan**[3], **Wendy K. Chung**[2,4], **Yufeng Shen**[1,5,6]

1 Department of Systems Biology, Columbia University, New York, NY, USA
2 Department of Pediatrics, Columbia University, New York, NY, USA
3 International Education College, Zhejiang University, Hangzhou, Zhejiang, China
4 Department of Medicine, Columbia University, New York, NY, USA
5 Department of Biomedical Informatics, Columbia University, New York, NY, USA
6 JP Sulzberger Columbia Genome Center, Columbia University, New York, NY, USA

Motivation
In-frame insertion and deletion variants (indels) alter protein sequence and length. Accurate predictions of pathogenicity are critically important for analysis of in-frame indels in genetic studies of human diseases. This is a challenging problem due to limitations in the available number of known pathogenic variants for training. Additionally, existing methods largely use supervised training with manually encoded features based on conservation, protein structure and function, and allele frequency. Recent advances in deep learning modeling of protein sequences and structures provide an opportunity to improve the representation of salient features based on large unlabeled protein sequences for prediction.

Results
We developed a new method (SHINE) for predicting pathogenicity of in-frame indels. SHINE uses pre-trained protein language models to construct a latent representation of an indel and its protein context from protein sequences and multiple protein sequence alignment, then feeds the latent representation to supervised machine learning models for classification tasks. We curated indels from ClinVar and gnomAD as positives and negatives, respectively, for training. To evaluate the performance and generalizability of the method, we created two test datasets from different sources. The first one includes de novo indels in large genetic studies of neurodevelopmental disorders (NDD) as positives. The second one includes indels in cancer mutational hotspots as positives. Short indels in the NDD genes and cancer mutational hotspot genes from large population genomes (UK biobank) served as negatives for the two test datasets respectively. SHINE achieved an area under the ROC curves of about 0.9 for both deletion and insertion variants in these two test data sets, which is significantly better than existing methods.

Conclusions
Our work suggests that unsupervised protein language models can provide valuable information for classification tasks in predicting pathogenicity of in-frame indels, and that the new method based on these models can improve interpretation of variants in genetic analysis.

Short-term effects of three personalized digital coaching prompts on physical activity

POSTER 280

Ali Javed[1], **Steven G Hershman**[1], **Marily Oppezzo**[1], **Anna Shcherbina**[1], **Donna M Rizzo**[2], **C Mikael Mattsson**[1], **Abby C King**[1], **Mathew T Wheeler**[1], **Euan A Ashley**[1]

1 Department of Medicine, Division of Cardiovascular Medicine, Stanford University, Stanford, CA 94305, USA
2 Department of Civil and Environmental Engineering, University of Vermont, Burlington, VT 05405, USA

Studies have shown that digital intervention using mHealth can increase short-term physical activity; however, to date, there has been no detailed investigation into the effects of personalized coaching prompts on physical activity. In this substudy of the MyHeart Counts study (ClinicalTrials.gov Identifier: NCT03090321), step count is used as a proxy for physical activity to study the short-term effects of three personalized coaching prompts: 1) prompt following one hour of inactivity, 2) daily inactivity prompt based on a threshold, and 3) daily tailored prompt based on baseline activity. A fourth category of daily prompt to read the American Heart Association website was used as a control. Baseline activity data were collected for one week prior to the start of the interventions (n = 2803, mean daily step count = 3776 ± 64). Tailored prompts were personalized to five activity pattern clusters created using a clustering algorithm on baseline data. Each participant was classified to one of the clusters and then randomly assigned to one of 24 permutations (four combinations of three 7-day interventions and the control) in a crossover design within a free-living setting. All interventions and the control resulted in a significant ($p<0.01$) increase in the mean daily step count from the baseline. The tailored prompt based on the baseline activity had the largest increase of 358 ± 61 daily steps ($p<0.001$, n = 2350), followed by the hourly inactivity prompt at 276 ± 61 ($p<0.001$, n = 2282). The daily inactivity prompt was not significantly different from the control module at an increase of 202 ± 62 ($p<0.01$, n = 2227) and 214 ± 62 ($p<0.01$, n = 2258) steps, respectively. However, we observed that participants classified to two of the clusters, labeled Weekend Warriors (n=189) and Drivers (n=84), responded to the daily inactivity prompt by a higher significant ($p<0.05$) increase of 595 ± 220 and 1067 ± 371 steps, respectively, but not to the tailored prompts or the control prompt. This study demonstrates that personalization is a feasible approach that can achieve more effective digital interventions for physical activity.

Spatiotemporal pattern of lymph node metastasis characterizes molecular features of esophageal squamous cell carcinoma

POSTER 285

Jiho Park[1], **Se-Young Jo**[1,2], **Jeongsoo Won**[1,2], **Seong Yong Park**[3,*], **and Sangwoo Kim**[1,2,*]

1 Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Korea
2 Graduate School of Medical Science, Brain Korea 21 Project, Yonsei University College of Medicine, Seoul, Korea
3 Department of Thoracic and Cardiovascular Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

Esophageal squamous cell carcinoma (ESCC) is a malignant neoplasm, accompanied by lymph node metastasis (LNM) to the neck, chest and abdomen. Although LNM is one of the most important prognostic factors, its progression patterns towards different anatomical sites are largely unknown. Here, we performed whole-exome (WES) and transcriptome sequencing (RNA-seq) of 47 multi-regional esophageal tissues (10 primary tumors, 10 normal esophagus and 27 lymph node tumors) from 10 ESCC patients, to reconstruct the spatio-temporal LNM patterns and profile genomic characteristics. Patients showed diverse LNM patterns in terms of the time to initial invasion, anatomical branching points, and genes mutated before and during metastasis. Somatic mutations acquired during LNM were largely sporadic in genes, but commonly presented two mutational signatures: defective DNA mismatch repair (SBS6) and damage by reactive oxygen species (ROS) (SBS18), suggesting distinctive mutational process in tumorigenesis. Also, tumors at later metastasized nodes acquired somatic mutations with a higher rate ($rs = 0.742$, $p<0.0001$) and can be more aggressive henceforth. Finally, we frequently observed (4/9, 44%) LNM that directly proceeded to anatomically distal lymph nodes in the current staging system, also known as nodal skip metastasis (NSM). Gene expression analysis of tumors with NSM revealed significant enrichment in epithelial-mesenchymal transition (EMT) (FDR < 0.001) and KRAS signaling (FDR<0.05), posing a further stratification strategy by LNM pattern. Overall, our study provides molecular understandings and potential usages of LNM patterns in staging ESCC.

Systematic evaluation of genome sequencing for the assessment of fetal structural anomalies

POSTER 290

Elise Valkanas[1-4,*], **Chelsea Lowther**[1-3,*], **Jessica L. Giordano**[5], **Harold Z. Wang**[1-2], **Benjamin B. Currall**[1-3], **Kathryn O'Keefe**[1-2], **Emma Pierce-Hoffman**[2], **Nehir Kurtas**[1-3], **Christopher W. Whelan**[2], **Stephanie P. Hao**[1,2], **Ben Weisburd**[1,2], **Vahid Jalili**[2], **Jack Fu**[1-3], **Isaac Wong**[1-2], **Ryan L. Collins**[1-2,6], **Xuefang Zhao**[1-3], **Christina A. Austin-Tse**[1,7], **Emily Evangelista**[2], **Gabrielle Lemire**[2], **Vimla S. Aggarwal**[8], **Diane Lucente**[1], **Laura D. Gauthier**[2,9], **Charlotte Tolonen**[2,9], **Nareh Sahakian**[2,9], **Christine Stevens**[2], **Joon-Yong An**[10], **Shan Dong**[11], **Mary E. Norton**[12], **Tippi MacKenzie**[12], **Bernie Devlin**[13], **Kelly Gilmore**[14], **Bradford Powell**[15], **Alicia Brandt**[15], **Francesco Vetrini**[16], **Michelle DiVito**[5], **Stephan J. Sanders**[11], **Daniel G. MacArthur**[1-2,18-20], **Jennelle C. Hodge**[16], **Anne H. O'Donnell-Luria**[1-2,17], **Heidi L. Rehm**[1-2], **Neeta L. Vora**[14], **Brynn Levy**[8,§], **Harrison Brand**[1-3,§], **Ronald J. Wapner**[5,§], **Michael E. Talkowski**[1-3,§]

1 Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA;
2 Program in Medical and Population Genetics, The Broad Institute of M.I.T. and Harvard, Cambridge, MA, USA;
3 Department of Neurology, Harvard Medical School, Boston, MA, USA;
4 Program in Biological and Biomedical Sciences, Division of Medical Sciences, Harvard Medical School, Boston, MA, USA;
5 Department of Obstetrics & Gynecology, Columbia University Medical Center, New York, NY, USA;
6 Program in Bioinformatics and Integrative Genomics, Division of Medical Sciences, Harvard Medical School, Boston, MA, USA;
7 Department of Pathology, Harvard Medical School, Boston, MA, USA;
8 Department of Pathology and Cell Biology, Columbia University Medical Center, New York, NY, USA;
9 Data Science Platform, Broad Institute of MIT and Harvard, Cambridge, MA, USA;
10 School of Biosystem and Biomedical Science, Korea University, Seoul, South Korea;
11 Department of Psychiatry, UCSF Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, USA;
12 Center for Maternal-Fetal Precision Medicine, University of California, San Francisco, USA;
13 Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA;
14 Department of Obstetrics and Gynecology, Division of Maternal-Fetal Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA;
15 Department of Genetics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA;
16 Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA;
17 Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA, USA;
18 Department of Medicine, Harvard Medical School, Boston, MA, USA; 19. Centre for Population Genomics, Garvan Institute of Medical Research, and University of New South Wales Sydney, Sydney, Australia;
20 Centre for Population Genomics, Murdoch Children's Research Institute, Melbourne, Australia.
* These authors contributed equally;
§ Co-senior authors.

CONTINUED

POSTER 290

Current clinical guidelines recommend three genetic tests to evaluate all variant classes for the assessment of fetal structural anomalies, but advances in whole-genome sequencing (WGS) and analysis suggest it is poised to become a single diagnostic approach and displace the sequential application of karyotype, chromosomal microarray (CMA), and whole exome sequencing (WES). To systematically benchmark the diagnostic yield of WGS, we developed an analytic framework that included the discovery, filtration, and interpretation of nine classes of genomic variation and applied it to 7,195 individuals. We assessed the sensitivity of WGS to detect diagnostic variants from three standard-of-care tests using 1,612 autism spectrum disorder (ASD) quartets with matched WGS, WES, and CMA then validated these findings in 46 fetuses with a clinically reportable variant originally identified by karyotype, CMA, or WES. We then assessed the added diagnostic yield of WGS in 249 trios comprising a prescreened fetus with a structural anomaly detected by ultrasound. Across these cohorts, WGS identified all 149 diagnostic variants detected by CMA and WES, as well as 11/14 of the balanced chromosomal rearrangements identified by karyotype. The diagnostic yield from WGS was 7.5% across all 1,612 ASD probands, almost two-fold more than both CMA (4.4%) and WES (3·0%). We also demonstrated that when including CNVs from exome data, the yield of WES (7.3%) can approach that of WGS. In 249 pre-screened fetuses with structural anomalies, WGS provided an additional diagnostic yield of 2.0% beyond the combination of karyotype, CMA, and WES. Applying our results to existing data indicates that WGS can achieve an overall diagnostic yield of 46.3% in unselected fetal structural anomaly cases. Significantly higher than karyotype (+14.4%), CMA (+13.9%), and WES (+38.5%). We demonstrate that WGS is sensitive to the detection of most pathogenic variation captured by karyotype, CMA, and WES, provides a superior diagnostic yield over any individual test, and increases diagnostic yield beyond the combination of all three tests. We outline several novel strategies that could be immediately implemented to improve the diagnostic yield of WGS. Overall, these data suggest WGS warrants consideration as a first-tier diagnostic approach for fetal structural anomalies.

Systematic follow-up to redefine recurrence risk of de novo mutations leading to pregnancy loss, terminations, and perinatal death

POSTER 295

Peer Arts[1,2], **Thuong T. Ha**[1,3], **Alicia B. Byrne**[1,4], **Matilda R. Jackson**[1,5], **Milena Babic**[1], **Marlie S.B. Frank**[1], **Tristan S.E. Hardy**[6,7], **David Lawrence**[3], **John Toubia**[3], **Sarah L. King-Smith**[1], **Alan McGovern**[3], **Barbara Nutchey**[3], **Thessa Kroes**[8], **Mark Corbett**[8], **Genomic Autopsy Study Research Network, Jozef Gecz**[8], **Karin S. Kassahn**[4,9], **Rob King**[3], **Andreas W. Schreiber**[2,9], **Christopher P. Barnett**[2,10#], **Hamish S. Scott**[1,2,3,5,6,#]

1 Department of Genetics and Molecular Pathology, Centre for Cancer Biology, an alliance between SA Pathology and the University of South Australia, Adelaide Australia
2 Adelaide medical School, University of Adelaide, Adelaide, SA, Australia
3 ACRF Genomics Facility, Centre for Cancer Biology, An alliance between SA Pathology and the University of South Australia, Adelaide, Australia
4 Center for Mendelian Genomics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
5 Australian Genomics, Melbourne, VIC, Australia
6 Department of Genetics and Molecular Pathology, SA Pathology, Adelaide, Australia
7 Repromed, Dulwich, SA, Australia
8 Adelaide Medical School and Robinson Research Institute, University of Adelaide, Adelaide, South Australia, Australia
9 School of Biological Sciences, University of Adelaide, Adelaide, Australia
10 Paediatric and Reproductive Genetics Unit, South Australian Clinical Genetics Service, Women's and Children's Hospital, North Adelaide, Australia
co-senior author

Background: The Australian Genomic Autopsy Study investigates the genetics underlying fetal loss, pregnancy terminations due to severe congenital abnormalities, and perinatal death. From exome and genome sequencing in 200 families, we have selected 34 (likely) pathogenic variants and 19 candidate variants which were called de novo. Follow-up of these de novo mutations is not routinely performed in the healthcare setting and recurrence risk is clinically counselled at ~1%.

Aim: To systematically assess parental mosaicism for de novo variants identified by the Australian Genomic Autopsy Study to allow more accurate counselling on individual recurrence risk.

Methods: We performed phasing for 49 de novo variants using existing genomic sequencing data (N=17), Nanopore sequencing (N=29) or ddPCR (N=3) to identify the parental origin. In addition, we applied custom droplet digital PCR (ddPCR) assays to quantify the presence of the specific variant in DNA from parental blood (72), saliva (28) and sperm (5) samples.

Results: Similar to earlier reports in postnatal genomic disorders, 33/41 (80%) of autosomal de novo mutations were phased to the paternal allele. Follow-up by ddPCR revealed parental mosaicism for 5/50 (10%) variants tested to date, with two paternal blood samples showing an allele balance >1% (2% and 10%%, respectively). While the variant at 10% was not called by the GATK haplotypecaller, the mosaicism could be visualised using IGV. Analysis of respective sperm DNA indicated a further increase in variant allele frequency (3% and 20% respectively) and associated recurrence risk.

Conclusion: Autosomal de novo mutations underlying fetal or perinatal death occur on the paternal allele in 80% of cases, supporting the idea of routinely assessing DNA from sperm samples to evaluate recurrence risks. While only 5 fathers provided sperm for ddPCR analysis, two of these samples displayed an elevated allele balance compared to blood. We recommend systematic use of somatic variant callers, manual inspection in IGV and assessment of low-level mosaicism by ddPCR in blood and sperm to facilitate more accurate counselling following identification of de novo variants.

Team-based genetic consultation: An effective system of care for the delivery of precision oncology services

POSTER 300

Maren T Scheuner[1,2], **Barbara Lerner**[3], **Randall Coeshot**[1], **Alicia Gable**[4], **Alexis Huynh**[4], **Paloma Sales**[1], **Katherine Hoggatt**[1,2], **Alison Hamilton**[4,5], **Marcia Russell**[4,5]

1 San Francisco VA Health Care System
2 UCSF School of Medicine
3 Boston VA Health Care System
4 VA Greater Los Angeles Health Care System
5 David Geffen School of Medicine at UCLA

Objectives: Department of Veterans Affairs (VA) patients with genetics referrals are 1.5 times more likely to have multiple cancer screening and preventive procedures if they complete their genetic consultations, but only when completed with a traditional program staffed by a team of clinical geneticists and genetic counselors versus a nontraditional, centralized telehealth program staffed by genetic counselors working independently. We sought to understand the reasons for these differences in cancer screening and prevention uptake.

Methods: We reviewed a stratified, random sample of medical records of patients with cancer genetics referrals (142 records for each care model). We conducted semi-structured interviews with 15 genetics providers and 36 referring clinicians. We analyzed the annotated medical records and interview transcripts using a rapid assessment process. We characterized annotations as personalized recommendations (e.g., "begin colonoscopy at age 30 then every 1-2 years"), options for consideration (e.g., "consider bilateral mastectomy"), and generic messages (e.g., "refer to the American Cancer Society guideline", "follow up with your gastroenterologist").

Results: Cancer screening or prevention was documented in 80 traditional-service records (141 annotations) and 106 nontraditional-service records (143 annotations). Personalized recommendations comprised 69% (97/141) of annotations within the traditional-service records and only 30% (43/143) within the nontraditional-service records. Generic messages comprised 17% (24/141) of annotations in the traditional-service records and 51% (73/143) in nontraditional-service records. From interview data, we learned referring clinicians expected a breadth of services, including management and screening recommendations, and referring clinicians stated their role was to follow through on recommendations made. Under the traditional programs, geneticists formulated recommendations documented. Under the nontraditional service, scope of practice limited how the genetic counselors addressed cancer screening and prevention.

Conclusions: Personalized recommendations were typical of traditional-program records, whereas nontraditional-program records usually had generic messages. Compared with the nontraditional program, the traditional programs are more patient-centered and better meet expectations of referring clinicians, which might explain, in part, the differences in patient uptake of cancer screening and preventive procedures.

Impacts: As the demand for genetic services grows, team-based care should be promoted for a more patient-centered, coordinated, and effective system of care.

Temporal dynamics of the multi-omic response to endurance exercise training across tissue

POSTER 305

David Amar[1], **Nicole R. Gay**[1], **Pierre Jean-Beltran**[2], **MoTrPAC Study Group**[3]

1 Stanford University School of Medicine, Stanford, CA 94305
2 Broad Institute of MIT and Harvard, Cambridge, MA 02142
3 Molecular Transducers of Physical Activity Consortium,National Institutes of Health

Regular exercise promotes whole-body health and prevents disease, yet the underlying molecular mechanisms throughout a whole organism are incompletely understood. The Molecular Transducers of Physical Activity Consortium (MoTrPAC) profiled the temporal transcriptome, proteome, metabolome, lipidome, phosphoproteome, acetylproteome, ubiquitylproteome, epigenome, and immunome in whole blood, plasma, and 18 solid tissues in Rattus norvegicus over 8 weeks of endurance exercise training. The resulting data compendium encompasses 9466 assays across 19 tissues, 25 molecular platforms, and 4 training time points in young adult (6-month-old) male and female rats. We developed a computational pipeline for the analysis of these multi-dimensional complex experimental data. We utilized empirical Bayes algorithms for identifying FDR-adjusted clusters of analytes that manifest coordinated temporal differential abundance patterns. Our results demonstrated distinct patterns of tissue remodeling, with widespread regulation of immune, metabolism, heat shock stress response, and mitochondrial pathways. These patterns provide biological insights into the adaptive responses to endurance training over time. For example, exercise training induced heart remodeling via altered activity of the Mef2 family of transcription factors and tyrosine kinases. Translational analyses revealed changes that are consistent with human endurance and candidate mechanisms that link training adaptation to non-alcoholic fatty liver disease, inflammatory bowel disease, cardiovascular health, and tissue injury and recovery.
The data and analysis results presented in this study will serve as valuable resources for the broader community.

The establishment of the Massachusetts General Hospital Genomic Medicine Unit to support the integration of genomics into clinical care

POSTER 315

Eugene K. Wong[1,2], **Renée C. Pelletier**[1], **Leland E. Hull**[4], **Emma K. Folkerts**[1], **Kristen M. Shannon**[1,5], **Christina Austin-Tse**[1] **& Heidi L. Rehm**[1,3,4]

1 Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA
2 Cardiology Division, Massachusetts General Hospital, Boston, MA
3 Broad Institute of MIT and Harvard, Cambridge, MA
4 Department of Medicine, Massachusetts General Hospital, Boston, MA
5 MGH Cancer Center, Massachusetts General Hospital, Boston, MA

While genetic and genomic testing in medicine has been growing rapidly over the years, numerous barriers have hindered its efficient integration into routine clinical practice. The Massachusetts General Hospital (MGH) Genomic Medicine Unit (GMU) was established to identify and address barriers to genomic medicine through implementation of practical solutions. Here, we report on our experiences addressing four major barriers.

First, genetics-related questions are commonly encountered by primary care providers (PCPs), who may have limited expertise or time to address them effectively. To address this, we developed: 1) point of care guideline modules within the EHR for commonly encountered genetic topics by PCPs 2) an online modular genomic medicine CME course for non-genetics providers, and 3) a Genetics and Genomics e-consultation (eConsult) service to address patient-specific genetics questions in the EHR. In 13 months, 70% of eConsults were placed by PCPs. Upon EHR review, 74% of providers implemented eConsult recommendations in their patient care plans.

Second, given increasing demands for genetic risk assessment among asymptomatic patients, we established the MGH Preventive Genomics Clinic (PGC). Since Fall 2019, the PGC has received over 220 referrals and coordinated genetic testing for most patients (64%), with hereditary cancer risk being the most common indication (41%). Third, there is need for genetic counseling (GC) support across subspecialties, including those with lower genetics patient volumes. We provided GC support through a 'floating' model, whereby a centrally-hired genetic counselor and GCA would be deployed based on clinic need.

Finally, significant variability exists in genetic lab utilization across the institution. From provider interviews and expert review of lab policies, we consolidated labs into three categories: preferred, acceptable, and discouraged. This classification will guide the use of high quality labs that adhere to professional standards, such as ClinVar data submission, ultimately increasing consistency, efficiency, and quality of patient care.

In summary, the GMU has been instrumental in supporting the implementation of genomic medicine through multiple tactical approaches. Future projects to support precision medicine will involve exploring more scalable service delivery models across high volume subspecialties, accreditation of genetic counselors, and EHR integration and genomic data capture with preferred laboratories.

The landscape of regional missense intolerance quantified from 125,748 exomes

POSTER 320

Katherine R. Chao, **Lily Wang, Genome Aggregation Database Consortium, Anne H. O'Donnell-Luria, Heidi L. Rehm, Daniel G. MacArthur, Michael E. Talkowski, Grace Tiao, Konrad J. Karczewski, Mark J. Daly, Kaitlin E. Samocha**

Missense variants comprise the majority of coding variants in a given genome and are known to be major drivers of human disease. However, they are notoriously difficult to interpret, and most missense variants observed to date have unclear clinical significance. Building more precise tools that aid in missense variant interpretation will expand our understanding of this class of variation and improve the diagnostic rate for individuals with suspected genetic disease.

The missense badness, PolyPhen-2, and constraint (MPC) score was developed to incorporate information about both regional missense constraint (RMC) and variant-level metrics in missense variant deleteriousness prediction. However, the initial derivation of these metrics had limited resolution due to the size of the population reference used (ExAC). Here, we leverage the power of 125,478 exomes in gnomAD v2.1 to update the MPC and RMC metrics for broader application in association studies and variant interpretation. Major method refinements improve the model of expected missense variation and introduce per-base resolution of constrained region breakpoints. The updated metrics will be available in the gnomAD browser, and the underlying code is open-source.

Using the updated metrics, we discover that 3,655 canonical gene transcripts harbor regional differences in missense constraint, 52.3% of which had escaped previous detection. Genic regions predicted to be highly constrained (observed/expected missense fraction < 0.4) are found to align more closely with protein domains. Initial analyses reveal a 10-fold enrichment of de novo missense variants predicted to be highly deleterious (MPC ⩾ 3) in 37,488 individuals with neurodevelopmental disorders compared to unaffected individuals, and an increase in predicted deleteriousness percentile of pathogenic ClinVar missense variants in genes associated with phenotypes caused by haploinsufficiency (Wilcoxon $p = 0.001$).

The improved metrics described here provide increased resolution on regional missense depletion within genes. We expect that the updated metrics will aid in variant interpretation and gene discovery efforts. As population references continue to grow, we will continue to gain insight into the landscape of missense constraint, improve predictions of missense variant consequences, and increase our ability to reduce the duration of patients' diagnostic odysseys.

The multi-omic mitochondrial response to endurance training across multiple tissues in MoTrPAC

POSTER 325

Malene E. Lindholm[1], **David Amar**[1], **Pierre M. Jean Beltran**[2], **Nicole R. Gay**[1], **Archana Raja**[1], **Megan Ramaker**[3], **Simon Schenk**[4], **Matthew T. Wheeler**[1], **MoTrPAC Study Group**[5]

1 Stanford University School of Medicine, Stanford, CA 94305
2 Broad Institute of MIT and Harvard, Cambridge, MA 02142
3 Duke University, Durham, NC 27708
4 University of California San Diego, San Diego, CA 92093
5 Molecular Transducers of Physical Activity Consortium, National Institutes of Health

Mitochondria are plastic organelles with diverse cellular functions critical to whole-body homeostasis. While chronic endurance exercise training is well known to alter aspects of mitochondrial activity in skeletal muscle, heart, liver and white adipose tissues, the effects in other tissues and organs are largely unknown, and multi-omic mitochondrial changes have not been investigated. In this study, as part of the Molecular Transducers of Physical Activity Consortium (MoTrPAC) we map the longitudinal, multi-omic changes in mitochondrial analytes across 19 tissues in 6 month-old male and female rats endurance trained for 1, 2, 4 or 8 weeks. The largest mitochondria-associated differential response was observed in the adrenal gland, brown adipose tissue, colon, heart and skeletal muscle, while there was little to no response in the brain (hippocampus, hypothalamus and cortex), lung, small intestine and testes. Almost 50% of all mitochondria-associated transcripts were altered in the adrenal gland, with an overall downregulation of fatty acid oxidation pathways in females across the 8 week time course, and upregulation of the same transcripts after the first week of training in males, reflecting a convergence of the sex-specific mitochondrial function of the adrenal gland with training. The female-specific response in colon showed opposite regulation of key mitochondrial metabolic pathways compared to the adrenal gland. Sex differences in the mitochondrial response were also prominent in the white adipose tissue. Training induced multiple changes in mitochondrial protein acetylation in the liver, including several electron transport chain and tricarboxylic acid (TCA) cycle proteins, to reflect the altered energy turnover following exercise. Brown adipose tissue (BAT) showed delayed downregulation of mitochondria metabolic pathways only after 8 weeks of training, with a concomitant male-specific reduction in mitochondrial DNA abundance. The striated muscle tissues demonstrated a highly coordinated mitochondrial response to increase oxidative capacity, with the majority of changes occurring in protein abundance and post-translational modifications. This comprehensive, multi-omic map of the temporal and sex-specific mitochondrial responses to training provides unique insights into previously unexplored tissues, with implications for mitochondrial function and health.

The reclassification odyssey: The research journey of genes and variants of uncertain significance to attain clinical relevance and prevent recurrent pregnancy loss

POSTER 330

Thuong T Ha[1,2], **Peer Arts**[1], **Alicia B Bryne**[1,3], **Marlie Frank**[1], **Jinghua Feng**[1,2], **David M Lawrence**[1,2], **John Toubia**[1,2], **Paul Wang**[1,2], **Milena Babic**[1], **Lynn Pais**[3], **Luis Arriola**[1,2], **Alison Gardner**[1,6], **Genomic Autopsy Study Research Network, Sarah L King Smith**[1], **Matilda R Jackson**[1,4], **Andreas W Schreiber**[2,5], **Anne O'Donnell-Luria**[3], **Christopher P Barnett6,**[7#], **Hamish S Scott**[1,2,4,6#]

1 Department of Genetics and Molecular Pathology, Centre for Cancer Biology, an alliance between SA Pathology and the University of South Australia, Adelaide Australia
2 ACRF Genomics Facility, Centre for Cancer Biology, An alliance between SA Pathology and the University of South Australia, Adelaide, Australia
3 Center for Mendelian Genomics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
4 Australian Genomics, Melbourne, VIC, Australia
5 School of Biological Sciences, University of Adelaide, Adelaide, Australia
6 School of Medicine, University of Adelaide, Adelaide, SA, Australia
7 Paediatric and Reproductive Genetics Unit, South Australian Clinical Genetics Service, Women's and Children's Hospital, North Adelaide, Australia
co-senior author

Background: Our Australian Genomic Autopsy Study has revealed that putative candidate gene(s) can be detected in half of clinically unresolved cases of early miscarriage or perinatal death (n=121/254). Among the candidate disease genes, only half (60/121) can be used prenatally to assess recurrence risk, despite compelling evidence from animal models, gene constraints and expression data. The ACMG guidelines are heavily weighted towards existing genotype to postnatal phenotype delineations, which reduces capacity to interpret findings from prenatal cases with either novel in utero phenotypes, or severe and/or lethal forms of existing genomic disorders.

Aims: Assess the proportion of variants and/or genes of uncertain significance that are eligible for reclassification based on additional genetic or experimental evidence.

Method: Genes with novel phenotypes or disease associations were submitted to MatchMaker Exchange (MME) or communicated directly with expert disease curators. Where fetal or parental tissue was available, RNA or methylation analyses was performed for validation of predicted splicing aberrations and/or potential epi-signatures. In a small subset, tailored in vitro assays were performed to assess the functional consequence of novel variants in known disease genes. Additionally, four mouse models have been established to characterize novel recessive disorders.

Results: 5/60 of the exome findings were reclassified to (likely) pathogenic, mostly from RNA analyses, while another 2/60 findings were reclassified from a gene of uncertain significance to a moderate evidence disease gene, based on the identification of additional kindreds and/or experimental evidence. 10/60 findings had at least one phenotype match from MME, of which four are currently included in large cohort studies for further phenotype delineations.

Conclusion: Among the candidates that were reclassified, 5/7 were recessively inherited, which relays a 25% chance of recurrence, while 2/7 were de novo and carries ~1% risk. Where gene novelty exists, identification of additional kindreds with overlapping phenotypes, via MME or direct correspondence, enables reclassification the fastest (1-2 years turnaround). Experimental evidence, particularly those involving animal models are unsurprisingly laborious (5-6 years turnaround), but imperative for highly intolerant or pleiotropic genes. There are translational benefits of systematically following up on candidate disease genes, which extends patient care beyond disease gene discovery.

Ultra-rapid whole genome nanopore sequencing in a critical care setting

POSTER 335

Background
Genetic disease is a major contributor to critical care hospitalization, especially in younger patients. While early genetic diagnosis can guide clinical management, the turnaround time for whole genome based diagnostic testing has traditionally been measured in months. Recent programs in neonatal populations have reduced turnaround time into the range of days and shown that rapid genetic diagnosis enhances patient care and reduces healthcare costs. Yet, most decisions in critical care need to be made on hourly timescales.

Methods
We developed a whole genome sequencing approach designed to provide a genetic diagnosis within hours. Optimized highly parallel nanopore sequencing was coupled to a high-performance cloud compute system to implement near real-time basecalling and alignment followed by accelerated central and graphics processor unit variant calling. A custom scheme for variant prioritization took only minutes to rank variants most likely to be deleterious allowing efficient manual review and classification according to American College of Medical Genetics and Genomics guidelines.

Results
We performed whole genome sequencing on 12 patients from the critical care units of Stanford hospitals. In 10 cases, the pipeline produced diagnostic results faster than all previously published clinical genome analyses. Per patient, DNA extraction, library preparation, and nanopore sequencing across 48 flow cells generated 173–236 GigaBases of sequencing data in as little as 1:50 hours. After optimization, the average turnaround time was 7:58 hours (range 7:18–9:0 hours). A pathogenic or likely pathogenic variant was identified in five out of 12 patients (42%). After Sanger or short read sequencing confirmation in a CLIA-approved laboratory, this validated diagnosis altered clinical management in every case.

Conclusions
We developed an approach to make a genetic diagnosis from whole genome sequencing in hours, returning actionable, cost-saving diagnostic information on critical care timescales.

Who defines the "personal utility" of genetic and genomic testing? A systematic review

POSTER 340

Katherine W. Saylor[1]

1 Department of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania

Predictive genetic screening in unaffected individuals is becoming more accessible, but its relevance for advancing health equity is debated. Many current efforts to pilot population screening are funded by government institutions with clear obligations to serve those most in need. To justify the continuation or expansion of these programs, policymakers need to understand the potential value of genetic screening as well as its equity effects. Cost-utility analysis focuses on the utilitarian ideal of maximizing total benefits, however, it also has the potential to inform questions about the distribution of benefits essential to non-utilitarian theories of allocation. Using a model of BRCA1/BRCA2 screening in White and Black sub-populations, I address efficiency, equality, and equity questions relevant to utilitarian, process egalitarian, outcomes egalitarian, and sufficientarian views. I compare three screening uptake scenarios: equal uptake; higher uptake in White women reflecting utilization disparities; and targeted screening for Black women only. I report that all three scenarios meet typical cost-effectiveness thresholds. Higher genetic testing uptake among White women is the most cost-effective strategy ($7,879/QALY gained) and averts the most premature deaths. All screening strategies decrease (but do not eliminate) per-person lifetime healthcare spending disparities related to breast and ovarian cancer prevention and treatment. Additionally, screening reduces disparities in quality-adjusted life expectancy, but only if screening probability is at least equal for Black women. Population BRCA1/BRCA2 screening, therefore, has the potential to improve health equity if programs ensure fair access to screening.

Whole genome sequencing improves the diagnostic yield of idiopathic dilated cardiomyopathy

POSTER 345

Katherine W. Saylor[1], **Elizabeth Q. Fernandes**[2], **Michael Adams**[3]

1 Department of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania
2 The University of North Carolina at Chapel Hill School of Medicine
3 Department of Pediatrics, Division of Pediatric Genetics and Metabolism, The University of North Carolina at Chapel Hill School of Medicine

Ovarian cancer (OC) is one of the leading causes of cancer mortality in women, accounting for 2.5% of cancers but 5% of cancer deaths in women (American Cancer Society 2019). The American Society of Clinical Oncology (ASCO) recommends germline testing for all ovarian cancer patients to guide treatment decisions (including the use of PARP inhibitors for women with BRCA1/BRCA2-related cancers) and cascade screening of family members. Yet, many patients do not receive timely germline testing. In a sample of 525 ovarian cancer patients diagnosed at the University of North Carolina in 2014-2019, we aimed to identify demographic and clinical factors associated with referral to genetics (325/525 = 62%), completion of germline testing among those referred (268/325 = 82%), and testing within 90 days among those tested after diagnosis (88/229 = 38%). We found substantial variation in referral rates across oncologists (ranging from 32% to 84% among oncologists who saw at least 15 patients in the sample). We modeled potential associations of referral and testing with patient race, age, insurance type, social deprivation by zip code, distance from the UNC hospital, oncologist, year of diagnosis, cancer stage, and tumor histology. When controlling for other variables, living more than 50 miles from the main UNC hospital was associated with lower odds of being referred to genetics compared to living within 20 miles (OR: 0.42, $p<0.009$). Medicaid insurance was associated with lower odds of testing once referred (OR: 0.23, $p=0.032$). More recent year of diagnosis was correlated with was associated with higher odds of timely testing after diagnosis (OR: 1.80, $p<0.001$). Our findings demonstrate some improvements in referral and testing rates over time but suggest a need for further exploration of uniform referral procedures within clinical practices, improved access to genetic testing for patients who are underinsured, and facilitation of remote testing options for patients who live far from genetics clinics.

Who defines the "personal utility" of genetic and genomic testing? A systematic review

POSTER 350

Emily G. Miller[1], **Jennifer L. Young**[1,2], **Anoushka Rao**[1], **Eliana Ward-Lev**[1] **and Meghan C. Halley**[1]

1 Stanford Center for Biomedical Ethics, Stanford University School of Medicine, Stanford, CA, USA
2 Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

Background: As clinical use of genetic testing grows, clinicians and researchers are increasingly recognizing that, beyond the clinical utility of these tests, there is substantial "personal utility" in receiving this information for patients and families. As our understanding of personal utility grows, it is essential to ensure that the voices of individuals with diverse sociodemographic backgrounds are included in defining and measuring this concept. We conducted a systematic review of peer-reviewed research focused on defining and measuring the personal utility of genetic and genomic testing to evaluate the underlying sociodemographic characteristics of the study participants.

Methods: Our systematic review utilized and updated a widely-cited 2017 systematic review by Kohler et al on the personal utility of genetic and genomic testing. Eligible studies reported empirical data on patient, family member, or public perspectives on the utility of any type of clinical genetic or genomic test. Individual-level demographic data were extracted, harmonized, and described using measures of central tendency and variance.

Results: Our review included 78 studies with 18,339 participants. Gender was the most frequently reported demographic characteristic (69 studies, 88%), followed by education (56 studies, 72%), race and ethnicity (49 studies, 63%), and income (35 studies, 45%).

Across all studies reporting, the majority of participants were female (61%, n=10,714), had a bachelor's or advanced college degree (62%, n=9,654), identified as non-Hispanic White (81%, n=9,335), and reported a household income above the U.S. median (64%, n=4,559). Across all categories, qualitative studies had an even greater percentage of participants in each of these categories, suggesting particularly limited sociodemographic diversity in research focused on defining the concept of personal utility.

Conclusions: Our results suggest that our understanding of the personal utility of genetics and genomics is disproportionately based on the perspectives of White women with high income and education. In order to achieve the National Human Genome Research Institute's commitment to equitable use of genomics in healthcare, we need not only to collect diverse genomic data but also to understand perspectives on the value of these data for diverse patients and families.

Whole genome sequencing improves the diagnostic yield of idiopathic dilated cardiomyopathy

POSTER 355

Seunghee Moon[1,¶], **Seung Hyun Lee**[2,¶], **Young Soo Chung**[1,¶], **Yoo-Jin Ha**[1], **Seok-Min Kang**[3], **Jaewon Oh**[3,*], **and Sangwoo Kim**[1,*]

1 Graduate School of Medical Science, Brain Korea 21 Project, Yonsei University College of Medicine, Seoul, Korea
2 Department of Biochemistry and Molecular Biology, Yonsei University College of Medicine, Seoul, Korea
3 Division of cardiology, Severance Cardiovascular Hospital, Cardiovascular Research Institute, Seoul, Korea
¶ These authors contributed equally.
* To whom the correspondence should be addressed.

Dilated Cardiomyopathy (DCM) is a fatal disease of the heart muscle with genetic penetrance, characterized by ventricular chamber enlargement and contractile dysfunction. Genetic testing has enabled the early diagnosis of familial DCM in up to 40%. However, only 15-25% of sporadic cases are genetically identifiable, demanding an urgent improvement. Here, we conducted Whole-Genome Sequencing (WGS) of 119 unrelated, panel-negative patients with idiopathic DCM to investigate the potential margin for improvement of diagnostic yield. We found that 15 patients (12.6%) can be newly diagnosable in WGS, exhibiting genetic variants suggestive of pathogenic (P) or likely pathogenic (LP), based on the currently recommended clinical guideline from the American College of Medical Genetics Guideline (ACMG). In particular, four (3.3%) had variants in coding regions that are not properly targetable, seven (5.8%) had cryptic splice variants, three (2.5%) had variants in promoter regions that disrupt gene expression, and one (0.8%) had a structural variation in exon region. We further identified 15 (12.6%) additional variants that are functionally interpretable to damage DCM causative genes, warranting an additional increase in diagnostic yield after validation or future amendment to the guideline. Our study poses an important augmentation of genetic diagnosis and emphasizes the role of WGS in conjunction with comprehensive bioinformatics analysis.

Accelerated computational pipeline for ultra-rapid nanopore whole-genome sequencing

POSTER 360

Whole genome sequencing can help in diagnosing genetic diseases, particularly in a critical care setting. But the existing sequencing and analysis pipelines limit the turnaround time for diagnosis. We developed an optimized nanopore-based sequencing approach to make genetic diagnosis from whole genome in less than 8 hours, returning actionable, cost-saving diagnostic information on critical care timescales. Specifically, we developed a high-performance cloud compute system to run in parallel and keep up with the rate of data generation from 48 flow cells. The system solved the challenges around real-time upload of the generated reads, efficient distribution of base calling and alignment compute across multiple GPU-enabled instances and minimization of associated inter-instance communication overhead that resulted in near real-time base calling and alignment. This was followed by accelerated GPU-based small variant calling and CPU-based structural variant calling parallelized across multiple cloud instances. As a result, our computational pipeline could generate annotated variant calls in almost 1hr 45min hours post-sequencing when we start with 212 giga bases of data.

Research Areas - Bioinformatics, whole genome sequencing, hardware acceleration, high performance cloud computing

Utilizing clinical natural language processing tools in the diagnosis of genetic diseases by rapid whole genome sequencing

POSTER 365

Edwin Juarez[1,2], **Kyu Seo Kim**[1,2], **Erica Sanford-Kobayashi**[1,2], **Christian Hansen**[1,2], **Lakshminarasimha Madhavrao**[1,2], **Christina Chambers**[3], **Stephen 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.

During the past decade we and others have developed methods for rapid diagnosis of genetic diseases by rapid whole genome sequencing (Dx-rWGS). In 31 published studies of these methods in critically ill infants and children in intensive care units (ICU), the rate of diagnosis was 37%, rate of acute change in management was 29%, and rate of change in outcome was 18%. In response to these reports, Dx-rWGS has started to be used more widely in children in ICUs in the US. There are ~7,150 known genetic diseases, and clinical presentations vary widely. Early ascertainment of infants in ICUs who will benefit from Dx-rWGS is crucial for optimal outcomes, but quite challenging because of genetic and clinical heterogeneity. This is exacerbated by broader use of Dx-rWGS by frontline neonatologists and intensivists who have little experience with this technology. We hypothesized that Electronic Health Records (EHRs), when combined with Clinical Natural Language Processing (CNLP), could be used to enrich for infants in ICUs who had undiagnosed genetic diseases. To test this hypothesis, we developed reproducible workflows to allow us to parse EHRs of 531 infants who had received Dx-rWGS, and identify key terms encoded as Human Phenotype Ontology (HPO) terms in order to identify terms that differ significantly in frequency in infants with and without genetic diseases. We optimized 2 CNLP tools and examined their analytic performance relative to manual extraction of HPO terms. We then processed the EHRs of 727 infants (including the 531 who received Dx-rWGS and 196 who did not, 425 who survived infancy and 302 who did not, 158 who received a diagnosis by Dx-rWGS and 373 who did not) using the NLP tools. Following quality checks, we compared the frequency of those terms across two conditions of interest (infant death vs infant survival, genetic disease diagnosis by rWGS vs undiagnosed by rWGS). We will show the results of univariate analyses and of predictive classifiers on both conditions. Such tools may be helpful in prioritizing infants in ICUs for Dx-rWGS.

3D genomics with Arima Hi-C sequencing enables detection of clinically relevant gene fusions in pediatric cancer samples

POSTER 500

Jon-Matthew Belton[3], **Kristin Sikkink**[3], **Midhat Farooqi**[1], **Atif Ahmed**[2], **Derek Reid**[3], **Anthony Schmitt**[3]

1 Children's Mercy Hospital Kansas City, Kansas City, MO 64108
2 Seattle Children's Hospital, Seattle, WA 98105
3 Arima Genomics Inc, San Diego, CA 92121

Gene fusions are known to have oncogenic effects in various cancers and serve as disease biomarkers. Obtaining gene-level resolution of these large structural variants is not possible with karyotyping and can be challenging using fluorescence in situ hybridization (FISH) or RNA sequencing due to low transcript abundance or low-quality RNA. Here, we used a novel method, Arima-HiC sequencing, on 7 archived pediatric cancer samples to determine the technique's effectiveness in detecting gene fusions. We first adapted the Arima Hi-C method for use with formalin-fixed paraffin-embedded (FFPE) tissue. We then selected five archived pediatric alveolar rhabdomyosarcoma (ARMS) tumors (FFPE archival period range: 8-12 years)—known to be fusion-positive via prior clinical testing. Briefly, FFPE tissue scrolls were dewaxed, and the tissue rehydrated, underwent chromatin fragmentation, end-labeling, and proximity ligation prior to DNA purification per the Arima-HiC FFPE protocol. Purified DNA was next prepared as a short-read sequencing library and sequenced on a HiSeq. The raw reads were aligned and deduplicated, and structural variants were called using the HiC-Breakfinder software. We additionally selected 2 non-FFPE pediatric leukemia cases with cryopreserved blasts (archival period range: 1-3 years), for Arima-HiC sequencing (as above). These cases had undergone standard of care cytogenetic (karyotyping, FISH) and molecular (targeted cancer NGS sequencing panel) testing clinically and a genetic driver / known gene fusion had not been identified. Using Arima-HiC sequencing in the first ARMS cohort, we identified either a PAX3-FOXO1 (n=3) or PAX7-FOXO1 (n=2) gene fusion in each of our 5 cases, consistent with the original diagnostic cytogenetic finding. HiC data was additionally able to provide previously unknown partner gene information (PAX3 or PAX7) in 4 of the 5 cases where the partner was previously not known. We then analyzed the cryopreserved blasts from the two leukemia cases without previously detectable genetic drivers or gene fusions, a precursor B-cell acute lymphoblastic leukemia (B-ALL) in a 3-year-old male, and acute myeloid leukemia (AML) in a 2-year-old female. In the first case, we detected an EP300-ZNF384 gene fusion, which is a known, but rare, gene fusion seen in B-ALL. In the second case, we detected a rare KMT2A-MLLT10 gene fusion. These findings are clinically important as both fusions are prognostically relevant. In summary, this study demonstrates how Arima-HiC sequencing provides molecular diagnostic value in archived pediatric solid and liquid tumor specimens via the identification of clinically relevant gene fusions.

A universal "day zero" infectious disease testing strategy exploiting CRISPR-based sample depletion and metagenomic sequencing

POSTER 505

The lack of preparedness for detecting the highly infectious SARS-CoV-2 pathogen — the pathogen responsible for the COVID-19 disease — caused enormous harm to the public health, the economy and society as a whole. It took ~60 days for the first RT-PCR tests for SARS-CoV-2 infection developed by the United States (US) Centers for Disease Control (CDC) to be made available. It then took >270 days to deploy 800,000 of these tests at a time when the estimated actual testing needs required over 6 million tests per day. Testing was therefore limited to only individuals with symptoms or individuals in close contact with confirmed positive cases. Testing strategies that can be deployed on a population scale at 'day zero' – i.e., at the time of the first reported case – are needed. Next Generation Sequencing (NGS) has day zero capabilities with the potential to enable feasible and very broad large-scale testing strategies. We show that the detection sensitivity of SARS-CoV-2 via NGS is equivalent to RT-PCR detection if relevant samples are depleted of abundant sequences that don't contribute to pathogen detection or host response. In addition, we show that the proposed strategy can also be used for variant strain typing, co-infection detection, and individual human host response assessment – all in a single workflow using existing open-source analysis pipelines. The NGS framework we describe is pathogen agnostic, and therefore has the potential to radically transform how both very large-scale pandemic response and focused clinical infectious disease testing are pursued in the future.

Adapting spatial transcriptomics protocols for archival FFPE specimens

POSTER 510

Radhakrishna Bettadapura, **Rashi Tomar, Jemima Jacob, Aditi Chatterjee, Aditi Gupta, Veena Ramaswamy, Ashish Kumar Choudhary, Yasodha Kannan Sivasamy, Vamsi Veeramachaneni, Arivusudar Everad John**

Spatial Transcriptomics (ST) can provide insights into positional expression of genes in tissues. The performance of Visium ST on fresh frozen tissues is relatively robust. However, its performance on archival FFPE samples in clinical settings is unclear. The goals of the study were to characterize the failure rates on FFPE samples, to assess relationship between IHC and gene expression, to understand the overlap between annotations made by pathologists and bioinformatics pipelines, and to assess if ST data provides greater insights than could be obtained through examination of H&E slides. The study was limited to surgical FFPE blocks <2 years old obtained from treatment-naive ER+ve breast cancer patients.

H&E slides were made from 64 FFPE blocks. Slides were assessed by a pathologist and blocks with low tumor content or excessive necrosis, stroma, fat were removed. Immune cell presence was noted but blocks were not removed based on this criteria. A section was taken from the remaining 45 FFPE blocks for RNA quality assessment. Only 25/45 blocks had DV200 above the Visium recommended threshold of 50%. Ten passed blocks and two with borderline DV200 values were shortlisted. Smaller FFPE sub-blocks were prepared to ensure that the final tissue specimen would fit within the boundaries of the capture area on the Visium slide.

Libraries were made from sub-block sections following Visium protocols. However, half the libraries had low yield and generated low-complexity data. 5 specimens generated good data which was analyzed using 10X Genomics software. The data was also analyzed using Seurat pipelines. In general, there is good overlap between spatial expression clusters generated by different pipelines indicating robustness of the underlying data. The clusters to a large extent correlate with pathologist annotations. Observations regarding cluster level marker genes, automated cell-type identification methods, pathology annotations, and IHC analysis will be presented.

Overall spatial data provided deeper insights into low-abundance cell types not identifiable by visual examination. However, laboratories need to be aware of the DV200 requirement, the sub-block preparation, and intricacies of library preparation to improve success rates.

Assessing the impact of applying ClinGen's recommendations for the use of calibrated computational prediction weights on missense variant classification

POSTER 515

Steven M. Harrison[1,2], **Alicia B. Byrne**[1]; **Heidi L. Rehm**[1,3]

1 Broad Institute of MIT and Harvard, Cambridge, MA
2 Ambry Genetics, Aliso Viejo, CA
3 Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA

Due to challenges in assessing pathogenicity, missense variants account for nearly 75% of variants of uncertain significance (VUS) in ClinVar. The 2015 ACMG/AMP guideline for sequence variant classification assigned deleterious missense predictions a "Supporting" level of evidence towards pathogenicity out of possible strength values VeryStrong, Strong, Moderate, and Supporting. A subgroup of the ClinGen Sequence Variant Interpretation working group calibrated missense predictors and found that multiple tools, including REVEL, can reach scores justifying the application of Moderate and Strong ACMG/AMP evidence criteria levels (https://www.biorxiv.org/content/10.1101/2022.03.17.484479v1).

To assess the impact of these recommendations, the calibrated weights of REVEL predictions were applied to all 2067 missense variants classified by ClinGen Variant Curation Expert Panels. This showed 26.1% (219/840) of current VUS missense variants could be upgraded to LP based solely on increased missense prediction weight. However, increasing the strength of missense predictors, while also including other predictive evidence (such as occurrence in a functional domain or a codon where other pathogenic variants have been observed), poses a risk of double-counting and over-classification. To account for this issue, classifications were re-assessed while limiting the combined total weight of all predictive data, including REVEL scores, at "Strong". With this approach, 18.7% (157/840) of current VUSs would reach a LP classification, with 100% (157/157) having additional supporting clinical or functional evidence applied. Comparatively, of the 62 VUS variants that no longer reach LP with a predictive data ceiling, only 25.8% (16/62) had additional supporting clinical or functional evidence applied. Additionally, when applying the calibrated missense predictor weights and limiting total predictive data strength, 0% (0/365) and 1% (5/478) of P and LP variants would be downgraded, respectively.

These findings demonstrate the potential of reducing VUS burden in clinical care through rigorous application of calibrated predictive tools.

Benchtop oligo synthesis: A faster and more reliable turnaround of assay results for genetic testing

POSTER 520

Lesnick, J.[1], **Quistad, S.**[1], **Nigg, J.**[2], **Saleh, M.**[2], **Ngan C.**[3], **Maurya, R.**[3], **Tardieu, N.**[1], **Lachaize, H.**[1], **Martinez-Fierro, M.**[4], **Garza-Veloz, I.**[4], **Centeno-Ramirez, A.**[5], **Derrien B.**[1], **Romero S**[1], **Lourenco N.**[1], **Kevin M.**[1], **Lachaize H.**[1], **Appel M.**[1], **Blumenfeld M.**[1], **Peponnet C.**[1] **and Wei C.**[3], **Godron X.**[1]

1 DNA Script, Paris, France & South San Francisco, CA, USA
2 Institute Pasteur, Paris, France
3 Genome Technologies at The Jackson Laboratory for Genomic Medicine, Bar Harbor, ME
4 Molecular Medicine Laboratory, Zacatecas, Mexico
5 Public Health Laboratory, Zacatecas, Mexico

Precision health is a set of scientific approaches and technical advancements that take into consideration an individual's genetic profile, environment, and lifestyle to determine a specific course of therapy. Characterization of the genetic profile of an individual requires the use of synthetic DNA oligos, typically supplied by external service providers, as primers for next-generation sequencing (NGS), Sanger sequencing, qPCR or FISH assays.

When demand for synthetic oligos is much higher than the supply available from service providers, such as during a pandemic, or when orders are delayed for logistical reasons, the turnaround time of the results from these assays becomes unpredictable and often delayed. Use of an in-house benchtop instrument to enzymatically synthesize oligos on demand can provide laboratories with more control over their workflows and a faster, more reliable turnaround time for results. Quick, on-demand and reliable access to oligos can provide researchers with the information they need to guide therapeutic decisions, when they need it.

In this poster we present and discuss the data resulting from the use of enzymatically synthesized primers and probes in a range of applications including amplicon-based targeted sequencing of a cancer panel, Sanger sequencing to confirm successful site-directed mutagenesis, RT-qPCR analysis of SMYD3 gene expression in HeLa cells, and single molecule RNA FISH (smFISH) as a powerful tool for the detection of viral RNA in the Drosophila gut.

BioModx: Accelerating molecular biology development through modular design and increasing the ease of deployment

POSTER 525

Symon Cotton[1], **Zachary Cater**[1], **Ben Hopkins**[1], **Matthew Armean-Jones**[1], **Carmen To**[1], **Claudia Bazyk**[1], **Eve Sherratt-Cross**[1], **Hazim Zameel**[1], **Holly Martin**[1]

1 Cambridge Consultants

BioModx is a unique modular system for automating molecular biology development, disrupting the conventional need for bespoke equipment. Through the use of rapidly 3D printed parts and the interoperability between BioModx units, this also provides a unique opportunity to deploy tailored diagnostic systems to developing countries and regions undergoing particular medical emergencies.

Most molecular biology protocols are underpinned by a combination of five or six discrete actions. Currently, custom developed diagnostic solutions combine and automate a specific sequence of these actions determined by specific protocols. BioModx changes this by separating each of these actions into a discrete block, enabling a rapid to deploy system that can be reconfigured subject to the protocol requirements.

In its first deployment BioModx has been developed to enable rapid diagnostic testing using loop-mediated isothermal amplification (LAMP) to test for conditions such as pneumonia. It uses a color-changing master mix to enable a distinctive result to be displayed, offering enhanced usability. In addition, diagnostic testing can be multiplexed running positive and negative control tests in parallel to two sample tests, or more depending on the microfluidic chip selected, with tests being approximately 20ul in volume.

Formed from off the shelf parts and 3D printable components, the supply of components has been simplified for the device and up-front tooling costs mitigated, reducing overall unit costs. In turn, this improves the accessibility of the device to a greater number of organisations.

Aid organisations are a strongly applicable audience that can use BioModx for rapidly deployed diagnostic testing in refugee camps. Due to these being geographically dispersed, with varying levels of infrastructure and personnel available, they are an ideal user group to ensure maximal accessibility to molecular diagnostics.

Developing automated diagnostic solutions requires significant investment, including the development of bespoke equipment needed to prove the validity of the approach. This bespoke equipment, however, minimises upfront cost whilst also extending the flexibility of the approach, should additional molecular biology protocols wish to be considered.

Characterizing ligation and amplification artifacts to ensure robust data analysis in a genome-wide integration assay assessing rAAVHSC editing

POSTER 530

Prout J[1], **Von Stetina J**[1], **Thompson J**[1], **and Wright J**[1]

1 Homology Medicines, Inc; Bedford, MA USA

For all gene-editing approaches, DNA changes need to be confirmed to determine whether any unintended changes have been introduced. rAAVHSCs have the advantage of integrating a transgene via homologous recombination (HR) into a specific location, but the presence of high levels of episomes makes assay design difficult. Viral integration studies generally include specific amplification from sequences on either side of the integration junction. Typically, primers directed at viral DNA are combined with primers directed at adaptors ligated to genomic DNA. The high degree of sample processing can lead to artifacts, which is problematic when identifying rare off-target integrations where false positives can obscure integration signals. At high doses, episomal rAAV may be at thousands of copies/cell and thus amplified or ligated inadvertently.

To overcome high rAAV episomal background, we modified previous protocols to minimize the impact of competing vector genomes. In addition, long-read sequencing ensures that viral and human genomic sequences are present on the same molecule and not inadvertently linked. Control experiments were carried out to ensure that our genome-wide integration assay is sensitive enough to detect low-level integration events and ensure that reads arising from artifactual ligations/amplifications are not mistaken for integrated reads. First, a limit-of-detection experiment was performed where a positive-control cell line with an engineered integration event was spiked into genomic DNA from mice transduced with a rAAVHSC vector at two doses. Next, a level-of-blank experiment was completed. The assay was performed on mice transduced with a rAAVHSC vector similar to the integrating rAAVHSC, but which had no homology arms, eliminating HR-based integration. Any positives from this experiment must have arisen from artifacts rather than true integrations, so they can be used to assess potential artifacts. Finally, simulated off-target plasmids were spiked into the positive-control cell line and the assay performed to ensure that a variety of sequence types could be detected. These controls assess assay artifacts arising from ligations or inappropriate amplification and potentially allow them to be removed via assay or data analysis changes. These controls are essential to ensure a robust integration assay that can be used to characterize targeted integration events using AAVHSCs.

Closing the precision medicine gap in immunotherapy through Predictive Immune Modeling

POSTER 535

As immuno-oncology continues to play an increasingly important role in cancer treatment, better diagnostics are needed to predict which patients will benefit from immunotherapies. A promising approach is characterizing the tumor-immune microenvironment, but current methods are limited. Single-analyte immunohistochemistry (IHC) approaches are widely used but have been a frustration in the industry and often show false positive rates as high as 90%. Flow cytometry accurately assesses the composition of infiltrating immune cells, but is incompatible with the formalin-fixed, paraffin-embedded (FFPE) specimens commonly used in the clinic. The key to bringing precision medicine to immunotherapy is multianalyte biomarkers that comprehensively and quantitatively characterize the immune component of the tumor microenvironment using novel approaches in accordance with clinical practice. Here we describe the first results coming out of PREDAPT, a national, multi-institutional study using Predictive Immune Modeling to determine which patients ultimately benefit from immunotherapy across more than 16 cancers approved for immune checkpoint inhibitors.

Using a cutting-edge approach to immune profiling of clinical FFPE tumor samples, we built transcriptome models for eight immune cell types as well as five T cell states. These Health Expression Models power predictive diagnostics compatible with sample format and quantity typical of clinical practice. The approach, Predictive Immune Modeling, outperformed PD-L1 IHC and Tumor Mutational Burden across three different cancer types (HNC, Melanoma, and NSCLC), nearly doubling accuracy compared to PD-L1 (from approximately 45% to 80%). Our results demonstrate that moving from single analyte to more comprehensive, multi-analyte immune biomarkers drives more accurate prediction of response to immunotherapy. Not only can Predictive Immune Modeling provide cost savings to the entire healthcare system, ultimately these advanced diagnostic tools lead to better treatment decisions and outcomes for patients.

Copy number alteration detection by targeted next generation sequencing in more than 12,000 real-world samples

POSTER 540

Grant Hogg[1], **Qian Zeng**[3], **Heidi Hoffmann**[3], **Eric A Severson**[2], **Li Cai**[2], **Stuart Schwartz**[2], **Wenjie Chen**[3], **Sabrina Gardner**[2], **Kimberly A Holden**[1], **Lax Iyer**[3], **Angela Kenyon**[3], **Deborah Boles**[2], **Scott Parker**[2], **Stanley Letovsky**[3], **Henry Dong**[4], **Narasimhan Nagan**[3], **Marcia Eisenberg**[5], **Anjen Chenn**[2], **Taylor J Jensen**[1,2]

1 Labcorp, San Diego, CA.
2 Labcorp, Durham, NC.
3 Labcorp, Westborough, MA.
4 Labcorp, New York, NY.
5 Labcorp, Burlington, NC.

In myeloid neoplasms cytogenetic and molecular analyses are standard of care for initial clinical evaluation. Next generation sequencing (NGS) is routinely used to identify somatic single nucleotide variants and insertions/deletions (<50 bp) (SNV/Indels) in clinically relevant genes to direct treatments, determine prognosis, and confirm diagnoses. Here we investigate whether data from a targeted NGS panel for SNVs/Indels can be utilized to determine copy number alterations (CNAs) also.

Real-world data from a targeted NGS panel (IntelliGEN® Myeloid) were retrospectively analyzed for research purposes to identify CNAs. The NGS panel was initially developed, validated, and implemented as a laboratory developed test to detect SNVs/Indels in 50 genes and CNAs in KMT2A. A preliminary cohort of 12,628 clinical samples from patients with cause-for-testing for myeloid malignancies including acute myeloid leukemia (AML), myelodysplastic syndrome (MDS), and myeloproliferative neoplasms (MPN) were re-analyzed for CNAs in 12 clinically relevant genes (BCOR, CBL, CDKN2A, CUX1, IKZF1, KMT2A, NF1, RUNX1, TET2, TP53, WT1, and ZRSR2). Copy number thresholds were set to ⩽0.85 for deletions and ⩾1.15 for gains with the goal of 99.7% specificity.

CNAs in these 12 genes were observed in 1607 (12.7%) samples, with a consistent rate over time. CNAs for 10 of the 12 genes were orthogonally confirmed in 377 samples by qPCR or digital multiplexed ligation dependent amplification (dMLPA) showing a confirmation rate of 89.3% (393/440). A significant correlation was observed between NGS and dMLPA (n=125, adjusted R2=0.92, p<.0001). Associations between small variants and CNAs detected in the same sample by NGS were also investigated. TP53 sequence mutations significantly co-occurred with CNAs in 11 of the 12 CNA genes (odds ratio = 2.8 to 39.9, adjusted p < 0.0004, all pairs) consistent with the known association between TP53 mutations and complex karyotypes.

These data suggest that a targeted NGS panel for SNVs/indels can also detect somatic CNAs. Further analysis is ongoing and these findings will be used to inform the development of improved NGS panels for somatic mutation and CNA detection. Potential applications of such a panel include monitoring genomic instability in response to treatment.

Detection of early-stage ovarian cancer using 5-hydroxymethylation profiles in plasma-derived cell-free DNA

POSTER 545

Anna Bergamaschi, **David Haan, Micah Collins, Gulfem Guler, Melissa Peters, Lauren Gigliotti, Shimul Chowdhury, Wayne Volkmuth, Samuel Levy**

Bluestar Genomics, San Diego CA, 92112

Background

Ovarian cancer is the fifth-leading cause of cancer-related deaths among women and accounts for more deaths than any other cancer of the female reproductive system. Ovarian cancer is often diagnosed in the advanced stage due to the lack of appropriate molecular tools that enable early detection.

We have developed a novel epigenomics detection methodology that enriches 5-hydroxymethylcytosine (5hmC) loci in cell free DNA (cfDNA) from plasma via click chemistry and sequencing. Our technology coupled with machine learning has enabled the generation of a classification algorithm capable of detecting cancers with high performance.

Method

Whole blood was obtained from a cohort of 888 individuals consisting of 92 ovarian cancers (OvCa) and 626 non-cancers. Plasma was processed to isolate cfDNA and 5hmC and low pass whole genome libraries were generated and sequenced. Logistic regression algorithms were employed using 5hmC feature sets from genes and genomic loci combined with physical characteristics of cfDNA (fragment size and copy number variation) to optimally partition cancer from non-cancer patients.

Results

To investigate whether OvCa can be detected in plasma, we interrogated plasma-derived cfDNA epigenomic signal derived from patients with ovarian cancer and non-cancer controls. We trained and cross-validated a binomial logistic regression model on OvCa and non-cancer samples utilizing both 5hmC features and physical characteristics of cfDNA, yielding an overall sensitivity of 83% (73%-90%), early-stage (stage I-II) sensitivity of 58% (28%-85%) and a specificity of 98% (96.8%-98.8%).

Conclusion

Our results demonstrate that plasma-derived cfDNA 5hmC profiles enable the detection of OvCa. Based on the performance in early-stage disease our technology lends itself to be a useful tool for the early detection of ovarian cancer especially for those individuals at highest risk for the disease. A larger clinical study is under development to validate the test in a high-risk population that would benefit most from early detection.

Developing precise genetic therapies for ultra-rare neurodevelopmental disorders: The case for WWOX

POSTER 550

Aqeilan R[1], **Berent A**[2], **Panagoulias J**[2], **Fyffe-Maricich S**[2], **Weiss Y**[2]

1 Hebrew University Jerusalem
2 Mahzi Therapeutics

Access to an abundance of genomic data and the ability to investigate it using advanced computational tools have led to identification of hundreds of genes associated with autism, developmental delays and seizure disorders (referred to as neurodevelopmental disorders, NDDs). The abundance of novel genetic disorders, along with parents and patients seeking various therapeutics approaches, necessitates the adoption of innovative approaches and collaborations towards patient diagnosis and therapeutic development.

WWOX (WW domain-containing Oxidoreductase) is one such gene. The WWOX gene codes for a scaffold protein that plays a role in the WNT, TGF-B and RTK signaling pathways. Homozygous recessive mutations in WWOX lead to severe epileptic encephalopathy presenting at infancy as WOREE Syndrome or SCAR12.

Proof of concept studies in an animal model of WOREE demonstrated that an AAV9-WWOX construct led to reversal of pathology including early mortality, failure to thrive and electrophysiological and behavior related pathologies (Repudi et al., EMBO Mol Med, 2021). Mahzi Therapeutics has in-licensed the program and is now scaling up vector manufacturing toward IND enabling studies and clinical trials, and working closely with the WWOX Foundation to develop a disease concept model, natural history database and ensure that all affected patients are diagnosed to help support a clinical development program for this rare, underserved disorder that has a tremendous clinical unmet need.

Efficient, scalable, and variant-robust targeted DNA and RNA capture using molecular inversion probes

POSTER 560

Eric D Boyden[1], **Jack M Amaral**[1], **Arjun D Patel**[1], **Joseph M Vieira**[1], **Gregory J Porreca**[1], **Patrick C Saunders**[1]

1 Molecular Loop Biosciences, Inc.

Common methods of target capture for next generation sequencing (NGS) include hybridization capture, which suffers from a complicated and labor-intensive workflow, and PCR capture, which is prone to poor uniformity and allele dropout. We have developed and optimized an alternative method based on molecular inversion probes (MIPs) that combines robust data quality with a highly scalable and easily automatable addition-only protocol that may be performed in a single workday. To demonstrate that neither the target size nor the number of MIPs is technologically limiting, we created and evaluated a full gene expanded carrier screening (ECS) panel that captures over 1 Mb of coding exons, splice sites, and 5' UTRs for 339 genes, using more than 43,000 tiled MIPs with an average of 5 MIPs per target base. The superior performance of this panel suggests that even full exome capture may be feasible, which would greatly improve on the simplicity and scalability of current methods without compromising data quality. Additionally, we developed an effective RNA capture methodology via the incorporation of reverse transcription into the MIP hybridization step. Unlike the PCR-based ARTIC assay, our SARS-CoV-2 (SC2) genome capture assay for short-read sequencing insures against amplification and mutation dropouts via capture of overlapping 225 bp elements at 7.5X mean tiling depth, thereby facilitating superior genome completion rates across a broad range of Ct values even for mutation-rich strains like Omicron. To establish that our chemistry can also produce libraries that are optimal for long-read sequencing, with greater probe redundancy and more power to identify indels and phase variants, we designed SC2 assays that capture overlapping elements of either 675 or 1215 bp, with mean tiling depths of 22.5X or 40.5X respectively. Their performance on synthetic RNAs shows not only that our long-fill SC2 genome capture assays are sensitive, specific, and robust against mutations, but also that MIPs can capture much larger elements than are commonly reported, which will improve their utility generally for long-read sequencing platforms.

Enzymatic Methyl-seq: Accurate methylation detection using picograms of DNA

POSTER 565

Vaishnavi Panchapakesa, **V. K. Chaitanya Ponnaluri, Daniel Evanich, Matthew A. Campbell, Romualdas K. Vaisvila, Bradley W. Langhorst, Eileen T. Dimalanta, Theodore B. Davis, Louise Williams**

New England Biolabs, Ipswich, MA 01938, USA

The epigenetic mechanisms that regulate gene expression, especially cytosine methylation, are important in understanding development as well as diseases such as cancer. Traditionally, methylation analysis of DNA samples with less than 10 ng input has been difficult primarily due to the use of bisulfite sequencing. DNA is severely damaged by sodium bisulfite treatment, resulting in regions of the genome with low coverage as well as biases introduced into the sequencing data. At lower DNA inputs these issues are harder to overcome even with increased sequencing depths. We have previously demonstrated that NEBNextÒ Enzymatic Methyl-seq (EM-seqÔ), an enzymatic alternative to bisulfite sequencing, gives robust and reliable methylation detection when compared to bisulfite sequencing for 10 to 200 ng of genomic DNA. DNA integrity was maintained with EM-seq libraries, which translated into reduced GC content bias and increased CpG detection compared to bisulfite sequencing.

Here we describe a method that reduces the input requirements for EM-seq down to 100 pg of genomic DNA. Low input Illumina compatible libraries were made using 100 pg to 10 ng NA12878 genomic DNA with the optimized low input EM-seq workflow. These libraries displayed similar metrics to standard input EM-seq libraries with even GC coverage distribution and high CpG coverage with strong correlation between replicates. The global methylation levels were in good agreement across the input levels from 100 pg to 10 ng. In addition, the number of CpGs covered were maintained at around 54-55 million at 1X coverage for 1 and 10 ng DNA and dropped to 25 million CpGs for 100 pg inputs. This low input method is robust and reproducible and gives researchers a new tool to tackle epigenetic studies where DNA has traditionally been limiting.

Estimating tumor fraction by fragmentomics in canine liquid biopsies

POSTER 570

Francesco Marass, **Kristina M. Kruglyak, Ilya Chorny**

PetDx, La Jolla, California, USA

Liquid biopsy has become a widely used tool in cancer research due to its ability to non-invasively interrogate tumor-derived cell-free DNA (cfDNA) repeatedly and at low risk to the patient, for applications such as cancer detection and targeted treatment selection. Liquid biopsy can also be used to measure tumor fraction in blood for prognostic and monitoring applications, for example after a therapeutic intervention or during treatment.

It has been well-documented that sequenced cfDNA displays a characteristic fragmentation pattern that carries information about the cellular processes that originate it. Features derived from this fragmentation profile have been used for cancer detection and have been shown to correlate with tumor stage in human patients.

Here, we show that the fragmentation profile can directly estimate the tumor fraction in canine blood samples by focusing on copy number-specific fragment length distributions of cancer-positive samples. Noting that each fragmentation profile is a mixture of a healthy profile and a tumor profile, we leverage the multiple fragmentation profiles per sample to deconvolute the signal from the two components. We thus obtain two pure fragmentomics profiles as well as an estimate of the sample's tumor fraction.

We developed our method on two sets of mixtures, where cfDNA samples with high tumor fraction were mixed in silico at predefined amounts into healthy cfDNA samples. Mixtures contained between 1% and 40% tumor-derived cfDNA. Validation was performed on two additional sets of mixtures. Overall, our predictions closely match expectations ($R2 > 0.94$), with estimated tumor fractions within 5% of true values. Performance improved with increasing tumor fraction. Finally, we analyzed a set of cancer-positive canine blood samples and show that the deconvoluted pure cancer profiles are shifted towards shorter fragments, in line with observations from genome-wide fragmentation patterns in humans. These results show that cfDNA fragmentomics can expand the clinical use of liquid biopsy tests to include quantitative monitoring applications in canine cancer patients.

Immune repertoire sequencing enables accurate clonality determination

POSTER 575

Chen Song, **Pingfang Liu, Ariel Erijman, Bradley W Langhorst, Andrew Barry, Salvatore Russello, Eileen T Dimalanta, and Theodore B Davis**

The study of complex immunological diseases and tumor microenvironments has advanced through recent developments in sequencing of the immune repertoire. Using this approach, the interrogation of disease progression is facilitated through analysis of millions of V(D)J combinations from B cell antibodies (BCRs) and T cell receptors (TCRs). One major challenge of immune repertoire sequencing is to accurately capture the structural and sequence complexities of antibodies and TCR genes during both library preparation and bioinformatic analysis. Here, we present a method for accurate sequencing of full-length immune gene repertoires of B cells and T cells.

RNA was extracted from tissues and peripheral blood mononuclear cells (PBMCs) and used for reverse transcription, during which unique molecular identifiers (UMIs) were added to discretely barcode each mRNA molecule. BCR- and TCR-specific PCR primers were used to enrich full-length BCR and TCR sequences. We have implemented a data analysis pipeline to assemble the full length BCR/TCR transcripts and to collapse PCR copies of each mRNA fragment into a single consensus sequence using UMIs. UMI incorporation enables the absolute quantification of input RNA molecules and accurate ranking of antibody/TCR clone abundance. Furthermore, this method facilitates detection of distinct and shared clones in tissue and blood samples, allowing identification of disease-specific clones to evaluate immunotherapy effects. Our method accurately and sensitively detects target TCR clones down to 0.01%, enabling minimal residual disease (MRD) assessment. Our immune repertoire sequencing approach allows accurate clonal determination for both BCR and TCR. This technique is applicable for a variety of applications including design of antibody chains for in vitro synthesis, investigation of T cell infiltration of tumor microenvironments, and monitoring of minimal residual disease in cancer patients.

Microsatellite instability detection with next-generation sequencing (NGS): A collaborative improvement utilizing clinical trial data

POSTER 580

Seifuddin Fayaz[1], **Ming-Shan Chien**[1], **Hung Ying Lin**[1], **Kevin Eng**[1], **Scott Chasalow**[1], **Ming Lei**[1], **Ryan Golhar**[1]

1 Bristol-Myers Squibb, Princeton, NJ 08648, USA

Context
Microsatellite Instability (MSI) has been reported as a predictive biomarker for checkpoint blockade in multiple tumor types. MSI can be tested by traditional PCR and lately using next- generation sequencing (NGS) data by computational methods.

Problem
Exploratory -omics data generated from clinical trials can vary due to differences in study designs, tumor type, intrinsic sample quality, pre-analytic sample handling, and downstream analytic protocols. Over the course of 21 BMS clinical trials between 2017 and 2021, we observed an unexpected increase in MSI-High (MSI Score >= 3.5) tumors assessed from Whole Exome Sequencing (WES) data using MSISensor 0.2. A collaborative team between Computational Genomics, Translational Bioinformatics Methodology, and Translational Medicine convened to understand the issue.

Methods
The team conjectured that this was an artifact resulting from NGS protocol changes that included increased depth of sequencing coverage. To test our hypothesis, we designed and executed a series of simulations to assess the validity of MSISensor's statistical test. We ran MANTIS and MSISensor2 as alternative MSI calling tools.R

Results
Our experiments show that MSISensor 0.2 is sensitive to depth of sequencing coverage, as depth increases, the MSI score gradually increases thus, calling more MSI-High samples. MANTIS and MSISensor2 are not sensitive to increase in depth of sequencing. In addition, the positive predictive value of MSISensor2 and MANTIS is higher than MSISensor 0.2. With respect to MSI-High samples, MSISensor2 and MANTIS are highly correlated. Compared to orthogonal MSI assays such as Biocartis and FoundationOne CDx, MSISensor2 under-calls MSI-High samples than MANTIS.

Conclusion
This study highlights the importance of paying careful attention to sequencing metrics and its effect on algorithms. In addition, this work demonstrates the importance of retrospective review to uncover artifacts not otherwise detectable in routine analysis. Consequently, we recommend using MANTIS in clinical NGS pipelines.

PointSuppressor: A rare allele detection technology devised for liquid biopsy–based cancer screening

POSTER 590

Charlie Rodi, **Saranya Rajavallaban**, **Venkat Ramamoorthi Elangovan**

RhoDx, Inc.

Early detection of cancer is considered key to good outcomes and when properly applied screening reduces cancer mortality. New systemic approaches detect circulating tumor DNA (ctDNA), found among cell-free DNAs (cfDNA) in the plasma of cancer patients. Detecting ctDNA in a systemic "Liquid Biopsy" has been used in cancer management, selecting patients for treatment, assessing drug efficacy, and monitoring for recurrence. In order to use the liquid biopsy to screen for cancer, high-throughput technologies and affordable approaches are of paramount importance in order to reach all segments of the at-risk general population and to test frequently enough to detect cancers early. Furthermore, the goal must be to detect cancer in as many individuals as practical. Data is presented from our analysis of the COSMIC database indicating that among solid tumors, 20 nucleotide positions found within 10 genes identify 75% of COSMIC entries among the top 100 mutated nucleotide positions in the human genome. Our PointSuppressor technology is ideally suited for detecting all of the corresponding mutations at these nucleotide positions. It is a single-tube method that dramatically reduces the amplification of wild type DNAs without interfering with the amplification of any mutation at a specific nucleotide position (e.g., a single assay detects all three mutations that occur at KRAS c.35 or any insertion or deletion initiating at the same nucleotide). It does so without extra processing, blockers, or sample loss, and does not introduce mutant sequences into the reaction as do technologies that use allele-specific PCR. Results demonstrate detection at all mutant frequencies tested, down to 0.07% mutant allele frequency using Sanger sequencing, 0.06% with MALDI-TOF MS, and 0.02% using Next Generation Sequencing (NGS). The suppression of uninformative wild type signals results in an approximate 1,000-fold combined benefit in sensitivity, throughput, and cost.

Estimating tumor fraction by fragmentomics in canine liquid biopsies

POSTER 595

Francesco Marass, **Kristina M. Kruglyak, Ilya Chorny**

PetDx, La Jolla, California, USA

Liquid biopsy has become a widely used tool in cancer research due to its ability to non-invasively interrogate tumor-derived cell-free DNA (cfDNA) repeatedly and at low risk to the patient, for applications such as cancer detection and targeted treatment selection. Liquid biopsy can also be used to measure tumor fraction in blood for prognostic and monitoring applications, for example after a therapeutic intervention or during treatment.

It has been well-documented that sequenced cfDNA displays a characteristic fragmentation pattern that carries information about the cellular processes that originate it. Features derived from this fragmentation profile have been used for cancer detection and have been shown to correlate with tumor stage in human patients.

Here, we show that the fragmentation profile can directly estimate the tumor fraction in canine blood samples by focusing on copy number-specific fragment length distributions of cancer-positive samples. Noting that each fragmentation profile is a mixture of a healthy profile and a tumor profile, we leverage the multiple fragmentation profiles per sample to deconvolute the signal from the two components. We thus obtain two pure fragmentomics profiles as well as an estimate of the sample's tumor fraction.

We developed our method on two sets of mixtures, where cfDNA samples with high tumor fraction were mixed in silico at predefined amounts into healthy cfDNA samples. Mixtures contained between 1% and 40% tumor-derived cfDNA. Validation was performed on two additional sets of mixtures. Overall, our predictions closely match expectations ($R2 > 0.94$), with estimated tumor fractions within 5% of true values. Performance improved with increasing tumor fraction. Finally, we analyzed a set of cancer-positive canine blood samples and show that the deconvoluted pure cancer profiles are shifted towards shorter fragments, in line with observations from genome-wide fragmentation patterns in humans. These results show that cfDNA fragmentomics can expand the clinical use of liquid biopsy tests to include quantitative monitoring applications in canine cancer patients.

Sensitive and unbiased detection of clinically actionable gene fusions from FFPE tumor biopsies using the Arima Hi-C platform

POSTER 600

Samir Sissaoui[1], **Michelle Perez-Arreola**[1], **David Jacob Hermel**[2], **Kristin Sikkink**[1], **Derek Reid**[1], **Anthony Schmitt**[1], **Darren S. Sigal**[3]

1 Arima Genomics Inc,
2 Scripps Clinic
3 Scripps MD Anderson Caner Center

Accurate and sensitive identification of gene fusions is critical to understanding disease mechanisms and the development and selection of optimal treatment regimens for cancer patients. Typically, these fusions are detected using low-resolution karyotyping, low throughput and biased FISH assays, or RNA sequencing approaches; however, the accuracy and sensitivity of gene fusion detection can be limited by factors such as low transcript abundance, transcript length, RNA degradation from formalin fixed paraffin embedded (FFPE) tissues, or the limited availability of fresh biopsy samples for RNA extraction. To address these limitations, we developed a novel approach to identifying gene fusions from FFPE samples using the Arima-HiC platform and short-read sequencing. We performed pan-cancer analysis on 12 FFPE adult tumor biopsies, each with gene fusions known to be clinically actionable. With this newly developed Hi-C workflow we identified all the known gene fusions with 100% sensitivity in the samples including those involving ALK, NTRK3, ROS1, FGFR2, and SS18 genes. This approach also revealed that the NTRK3 gene fusion was the result of a more complex rearrangement within chr12, within chr15, and between chr12 and chr15. Additionally, we detected the presence of numerous structural variants per sample in addition to the known gene fusion in each sample. For example in a bile duct tumor, Arima-HiC detected an FGFR2-EEA1 gene fusion, as well as 25 other structural variants genome-wide including in a CPD-LASP1 gene fusion This gene fusion has not been reported to our knowledge, however, LASP1 is a reported 3' fusion partner with MLL in leukemia. While these findings are not clinically actionable today, the unbiased, accurate, and sensitive detection of structural variants from frozen and FFPE solid tumor biopsies may facilitate further research into their relationship between disease mechanisms and clinical outcomes. Taken together, these findings demonstrate the analytical utility of Arima Hi-C sequencing technology to provide both chromosome-scale and gene-level resolution for the detection of structural variants in tumor biopsies samples. This workflow can provide improved access to critical genomic information from FFPE blocks for the identification of pathognomonic and druggable gene fusion events and other structural variants across tumor types.

Targeted detection of clinically relevant variants in HBA1/2 from whole genome sequencing data

POSTER 605

Shunhua Han[1], **Vitor Onuchic**[1], **Massimiliano Rossi**[1]

1 Illumina Inc, San Diego, CA

α-thalassemia is an inherited and autosomal recessive blood disorder caused by mutations in HBA1 and/or HBA2 genes that lead to reduced production of α-globin chains. It is also one of the most common human monogenic disorders. It has been estimated that approximately 5% of the world population are carriers of α-thalassemia variants [1], and the carrier frequencies in African population can be as high as 30-40% [2]. About 95% of α-thalassemia cases result from gene deletion(s) rather than non-deletional variants [3]. Targeted next-generation sequencing (NGS) technology has been demonstrated as a screening and confirmation method for α-thalassemia with high sensitivity [4,5]. However, there is a lack of approach to detect α-thalassemia variant from standard whole genome sequencing (WGS) data, in part due to high homology between HBA1 and HBA2 gene regions that results in ambiguous read alignment. Here we propose a novel computational approach for detecting both deletional and non-deletional variants of HBA1/2 from standard Illumina WGS data. This new approach allowed us to detect and genotype the common clinically relevant deletions in HBA1/2, including one copy deletions such as α3.7, α4.2, and two copy deletions in cis such as SEA. In addition, the approach also covers 17 pathogenetic/likely pathogenetic small variants with high review status in ClinVar.

We applied this new approach on 170 samples randomly selected from the 1000 Genomes (1kG) Project and compared the copy number genotypes with ones from orthogonal long read technology. This analysis showed that the positive percent and negative percent agreement between our approach and orthogonal technology are 100% (31/31) and 96% (134/139), respectively. We also applied our approach on 3201 samples with diverse genetic background from 1kG Project. Our approach successfully predicted copy number genotypes for 98.5% of samples and achieved 100% mendelian inheritance consistency for all 575 trios included in this dataset without missing genotypes. This analysis suggested that the carrier frequencies in the African and South Asian populations are approximately 36% (313/881) and 13% (77/593), respectively, and that two-copy deletion in cis has highest frequency in the Dai Chinese population (~14%, 13/90), which is consistent with previous findings [6]. Overall, our study demonstrates the potential to use WGS for both α-thalassemia carrier screening and diagnosis.

Uncovering gene fusions with 3D genomics: From clinical validation to actionable insights for undiagnosable solid tumors

POSTER 610

Anthony Schmitt[2], **Kristin Sikkink**[2], **Kristyn Galbraith**[1], **Michelle Perez-Arreola**[2], **Misha Movahed-Ezazi**[1], **George Jour1**, **Matija Snuderl**[1]

1 Dept of Pathology, NYU School of Medicine, New York, NY, 10016, USA
2 Arima Genomics Inc, San Diego, CA, 92121

Identifying gene fusions in tumor biopsies is critical for understanding disease etiology, however, clinical NGS panels often fail to yield clear genetic drivers. One challenge is that RNA-seq does not perform well in FFPE tissue blocks due to RNA degradation, low transcript abundance, and/or RNA panel design. To overcome this, we developed a novel DNA-based partner-agnostic approach for identifying fusions from FFPE tumors using 3D genomics based on Arima-HiC technology, in some cases with target enrichment (Capture-HiC), and NGS. Using this approach, we have profiled 108 FFPE tumors across tumor types. We first performed clinical validation of the Capture-HiC approach by re-analyzing 29 FFPE tumors comprising 34 actionable gene fusions detected by the RNA-based NYU FUSION SEQer CLIA assay. We observed a 91% concordance (31/34) between Capture-HiC and RNA panels, and are currently re-testing the 3 discordant fusions. We then analyzed 63 FFPE tumors using genome-wide HiC, including 36 CNS tumors, 10 gynecological sarcomas, and 12 solid hematological tumors (lymphoma / plasmacytoma). These tumors had no genetic drivers from prior CLIA-validated DNA and RNA panels. Amongst these, HiC analysis identified previously undetected fusions in 71% (45/63) of tumors. To attribute clinical significance to the fusions detected, we compared the genes implicated in our fusion calls with NCCN and WHO guidelines, and OncoKB, and assigned which tumors had a therapeutic level biomarker (e.g. PD-L1, NTRK, RAD51), or a diagnostic / prognostic biomarker (e.g. MYBL1 in glioma). We found 39.7% (25/63) of tumors had fusions involving a therapeutic level biomarker and a further 12.7% (8/63) had fusions involving a diagnostic or prognostic biomarker, indicating an overall diagnostic yield of 52.4%. The remaining 19% (12/63) had fusions of potential clinical significance, according to OncoKB. To highlight examples, we identified a novel PD-L1 rearrangement in a pediatric glioma that was not detected by DNA or RNA panels. Our finding was confirmed by PD-L1 IHC, and the patient was put on pembrolizumab off-label after tumor recurrence and has exhibited a complete response. We also identified MYBL1 fusions in two glioma cases that were previously missed by RNA panels. In one case, our MYBL1 fusion spared the patient unnecessary chemo post resection. Together, our findings demonstrate clinical validation, and highlights the utility for 3D genome profiling to increase diagnostic yield by finding clinically actionable fusions as a reflex test and as a frontline test in tumors without available NGS fusion assays (e.g. solid hematological tumors).

Whole genome sequencing–based homologous recombination deficiency testing for precision oncology of breast cancers

POSTER 615

Ryul Kim[1], **Seongyeol Park**[1], **Joonoh Lim**[1], **Boram Yi**[1], **Jaemo Koo**[1], **Sangmoon Lee**[1], **Jeong Seok Lee**[1,2], **Yeon Hee Park**[3], **Young Seok Ju**[1,2]

1 GENOME INSIGHT Inc, Daejeon, Republic of Korea
2 Graduate School of Medical Science and Engineering (GSMSE), Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea.
3 Division of Hematology-Oncology, Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

Breast cancers (BCs) with defects in homologous recombination repair pathway depend on the Poly (ADP-ribose) polymerase enzymes (PARP) for their survival and thus are sensitive to PARP inhibitors (PARPi). To screen out BCs with homologous recombination deficiency (HRD), germline mutations in BRCA genes have been tested, which constitute approximately 5% of unselected BC patients. However, a substantial proportion of BCs show HRD phenotypes despite the absence of BRCA inactivating mutations in germline. Herein, we apply whole-genome sequencing in 279 BC samples for characterizing their HRD phenotypes for PARPi treatments of 279 BC patients, 50 cases (17.9%) show remarkable HRD phenotypes, based on mutational signatures of HRR abrogation. Intriguingly, 23 cases do not harbor any germline mutations in HR-related genes including BRCA1 and BRCA2. Although 10 and 8 have somatic BRCA1 and BRCA2 mutations, respectively, 12 show no remarkable mutations in the pathway. Landscapes of somatic mutations in BCs are different according to HRD status. Overall, our analysis demonstrate that WGS enables better classification of the HR status in BC than conventional BRCA germline testing, suggesting WGS as a rapid and sensitive tool for screening breast cancers for PARPi treatment.

Index

Raja	Archana	Stanford University School of Medicine	55
Rajavallaban	Saranya	RhoDx, Inc.	80
Ramaker	Megan	Duke University	55
Rao	Anoushka	Stanford University School of Medicine	60
Rehm	Heidi L.	Broad Institute of MIT and Harvard Center for Genomic Medicine	47
Reid	Derek	Arima Genomics Inc	67, 85
Rehm	Heidi L.	Massachusetts General Hospital	51, 53, 54, 67
Relangi	Tulasi	University of Texas Southwestern Medical Center	22
Rhee	Yumie	Yonsei University College of Medicine	34
Rhie	Arang	National Institutes of Health	25
Rizzo	Donna M.	University of Vermont	45
Robbins	Sarah	National Institutes of Health	19
Rodi	Charlie	RhoDx, Inc.	80
Rocheleau	Ghislain	Icahn School of Medicine at Mount Sinai	40
Romero	S	DNA Script	68
Rosenson	Robert S.	Icahn School of Medicine at Mount Sinai	40
Rossi	Massimiliano	Illumina Inc	83
Russell	Marcia	VA Greater Los Angeles Health Care System David Geffen School of Medicine at UCLA	50
Russello	Salvatore		78

S

Sahakian	Nareh	The Broad Institute of M.I.T. and Harvard Broad Institute of MIT and Harvard	47
Saleh	M.	Institute Pasteur	68
Sales	Paloma	San Francisco VA Health Care System	32, 39, 51
Samocha	Kaitlin E.		54
Sanders	Stephan J.	University of California, San Francisco	47
Sanford-Kobayashi	Eria	Rady Children's Institute for Genomic Medicine Rady Children's Hospital	63
Sapp	Julie C.	Center for Precision Health Research National Human Genome Research Institute National Institutes of Health	42
Saunders	Patrick C	Molecular Loop Biosciences, Inc.	75
Saylor	Katherine W.	University of Pennsylvania	58, 59
Sedlazeck	Fritz J.	Baylor College of Medicine	25, 36
Sera	Tony Di	University of Utah School of Medicine	23
Schachter	Mike	Parker Institute for Cancer Immunotherapy	38
Schenk	Simon	University of California San Diego	55
Scheuner	Maren T	San Francisco VA Health Care System	32, 39, 51
Schmitt	Anthony	Arima Genomics Inc	64, 82, 84
Scholz	Sonja	National Institutes of Health	25
Schreiber	Andreas W	Centre for Cancer Biology, An alliance between SA Pathology and the University of South Australia University of Adelaide	49, 56
Schwartz	Stuart	Labcorp	72
Scott	Hamish S.	Centre for Cancer Biology, an alliance between SA Pathology and the University of South Australia Australian Genomics University of Adelaide SA Pathology	49, 56
Severson	Eric A	Labcorp	72
Serway	Christine N.	VA National Oncology Program Office	39
Shafin	Kishwar	Google LLC	25
Shannon	Kristen M.	Massachusetts General Hospital	53
Shcherbina	Anna	Stanford University	45
Shen	Yufeng	Columbia University	45
Sherratt-Cross	Eve	Cambridge Consultants	69
Shim	Hyo Sup	Yonsei University College of Medicine	17
Sigal	Darren S.	Scripps MD Anderson Cancer Center	82
Sivasamy	Yasodha Kannan		66
Sikkink	Kristin	Arima Genomics Inc	64, 82, 84
Sissaoui	Samir	Arima Genomics Inc	82
Skead	Kimberly	Ontario Institute for Cancer Research University of Toronto	43
Santulli-Marotto	Sandra	Parker Institute for Cancer Immunotherapy	38
Smith	K.S.	Stanford University	33
Smith	Sarah L King	Centre for Cancer Biology, an alliance between SA Pathology and the University of South Australia	49, 56
Snuderl	Matija	NYU School of Medicine	84
Soave	David	Wilfrid Laurier University	43
Song	Chen		74
Sousa	Nayara	UC San Diego	30
Stevens	Christine	The Broad Institute of M.I.T. and Harvard	47

T

Talkowski	Michael E.	Massachusetts General Hospital The Broad Institute of M.I.T. and Harvard Harvard Medical School	47, 54