AGBT 2022™

GENERAL MEETING

June 6 - 9, 2022 • Orlando, FL

SIGNIA BY HILTON • ORLANDO BONNET CREEK

AGBT 2022™
GENERAL MEETING

June 6 - 9, 2022 • Orlando, FL

It is our honor to welcome you to the **Advances in Genome Biology and Technology (AGBT) General Meeting 2022**. For many of us, this is the first time we are gathering in-person since AGBT 2020. It is good to be back!

Now in its 22nd year, AGBT is recognized as a premier gathering for the genomic research community.

Over the next few days, we will focus on collaboration and dialogue regarding the latest advances in DNA sequencing technologies, experimental and analytical approaches for genomic studies, and their myriad applications.

AGBT is a not-for-profit organization providing three prestigious conference and networking events, including The General Meeting (June 6-9) AGBT Precision Health (Sept 8-10) and AGBT Ag (March 27-29, 2023).

Organizing Committee

We are incredibly grateful to our meeting Co-Chairs Eric Green, Len Pennacchio, Beth Shapiro, and the entire Scientific Organizing Committee for the countless hours they invested in making AGBT22 a content rich experience.

Penelope Bonnen
Baylor College of Medicine

Federica Di-Palma
Genome British Columbia, Vancouver CA / University of East Anglia, UK

Kim Doheny
Johns Hopkins University School of Medicine

Stacey Gabriel
Broad Institute of MIT and Harvard

Eric Green
National Human Genome Research Institute, Meeting Co-Chair

Martin Hirst
University of British Columbia

Christopher Mason
Weill Cornell Medicine

John McPherson
University of California, Davis

Stephen Montgomery
Stanford University School of Medicine

Len Pennacchio
Lawrence Berkeley National Laboratory, DOE Joint Genome Institute, Meeting Co-Chair

Beth Shapiro
University of California, Santa Cruz, Meeting Co-Chair

Michael Zody
New York Genome Center

AGBT 2022™
GENERAL MEETING

We gratefully acknowledge the generous support provided by the following companies

GOLD

resolve biosciences

SILVER

ULTIMA GENOMICS
Element Biosciences
10x GENOMICS

BRONZE

TWIST BIOSCIENCE
PacBio
DNASCRIPT
IDT INTEGRATED DNA TECHNOLOGIES
MIROCULUS
Agilent Trusted Answers
NEW ENGLAND BioLabs Inc.
BRUKER
bionano GENOMICS

EXHIBITORS AND CONTRIBUTING SPONSORS

Veranome Biosystems
nanoString
illumina
AKOYA BIOSCIENCES THE SPATIAL BIOLOGY COMPANY
REBUS BIOSYSTEMS
Volta Labs
dolomite bio
Spatial Genomics
BioSkryb GENOMICS
TWINSTRAND BIOSCIENCES
molecular loop
PerkinElmer For the Better
imec embracing a better life
Cantata Bio
Miltenyi Biotec
BD
CODEXIS
MGI
IONPATH
QIAGEN
BIO-RAD
WATCHMAKER GENOMICS
Roche
vizgen
LEVITAS BIO™
TECAN
SINGULAR GENOMICS
PGDx labcorp

General Information

Registration / Hospitality Desk

Sunday,	June 5,	11:30 am – 7:00 pm
Monday,	June 6,	8:00 am – 7:00 pm
Tuesday,	June 7,	8:00 am – 7:00 pm
Wednesday,	June 8,	8:00 am – 7:00 pm
Thursday,	June 9,	8:00 am – 6:00 pm

Name Badges

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

Social Bands will also be available to offer a visible indicator of your personal comfort level in today's COVID environment.

Transportation

Orlando International Airport is 21 miles from the Signia by Hilton Orlando Bonnet Creek. The transfer time to the resort is approximately 25 minutes, depending on traffic. AGBT will offer complimentary shuttle service for arrivals and departures to and from the airport and hotel on Sunday, June 5 (1:00 PM – 8:00 PM), Monday, June 6 (8:00 AM – 8:00 PM), and Friday, June 10 (4:00 AM – 2:00 PM). Note: If you are arriving or departing on an alternate date, you will be responsible for securing your own transportation to the airport.

Social Events

AGBT Welcome Reception – **Monday, June 6 from 7:00 pm – 10:00 pm** at the Hilton Pool.

Women's Networking Event – **Tuesday, June 7 from 5:30 pm – 7:15 pm** on Signature Island.

Click2Win – **Tuesday, June 7 at 9:30 pm** in the Sponsor Promenade.
Make sure to visit the Sponsor Promenade to complete your Click2Win challenge. Attendees can earn CLICK2WIN points by completing different photo challenges to earn badges.

Grand Prize is One Complimentary 4-Day registration to GM 2023! The Grand Prize will be awarded during Thursday's Plenary Session. (This prize is non-transferable.)

AGBT Dinner – **Wednesday, June 8 from 6:15 pm – 7:25 pm** on the Outdoor Waldorf Promenade.

AGBT Farewell Dinner Party – **Thursday, June 9 from 6:00 pm – 10:00 pm** at the Golf Pavilion.

Conference Technology

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

Start Tweeting!

#AGBT22

ABSTRACTS DUE JUNE 16TH

EARLY BIRD REGISTRATION CLOSES JULY 14TH

LATE BREAKING ABSTRACTS DUE AUGUST 1ST

Agenda

Monday, June 6 — **AGBT Meeting and Pre-Conference Registration**

10:00 a.m. – 4:00 p.m. — **Pre-Conference**

10:00 a.m. – 11:00 a.m. — **McDonnell Genome Institute, Washington University**
Human Pangenome Reference Consortium (HPRC)
• Floridian Ballroom

Valerie Schneider, Ph.D., National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health
"Navigating the changing human genome resource landscape: GRCh38, T2T-CHM13 and HPRC assemblies"

Arang Rhie, Ph.D., National Human Genome Research Institute, National Institutes of Health
"A finished Y chromosome from HG002 completes the T2T-CHM13 reference"

Xian Chang, UC Santa Cruz Genomics Institute
"The HPRC human pangenome graph: updates and applications"

12:00 p.m. – 4:00 p.m. — **NanoString Spatial Multiomics Symposium**
• Bonnet Creek Ballroom

Christopher Mason, PhD, Professor; Director, WorldQuant Initiative, Weill Cornell Medicine

Joseph Beechem, PhD, Chief Scientific Officer, Sr. Vice President, R&D, NanoString Technologies

Grant R. Kolar, M.D., PhD, Associate Research Professor, Saint Louis University School of Medicine

Arutha Kulasinghe, PhD, NHMRC Research Fellow, The University of Queensland

Kevin Wei, M.D., PhD, Assistant Professor of Medicine, Brigham and Women's Hospital

Miranda Orr, Ph.D., Assistant Professor, Gerontology and Geriatric Medicine, Wake Forest University School of Medicine

Scientific Program

8:00 a.m. – 7:00 p.m.	**Meeting Registration / Hospitality Desk** • Bonnet Creek Ballroom Center Foyer
Plenary Session:	**Genomics I (Eric Green, National Human Genome Research Institute, Chair)** • Floridian Ballroom Salon A-L
5:00 p.m. – 5:05 p.m.	**Opening Remarks**
5:05 p.m. – 5:35 p.m.	Euan Ashley, Stanford University *"Long read genome sequencing: speeding towards the clinic"*
5:35 p.m. – 6:05 p.m.	Olga Troyanskaya, Princeton University, and Deputy Director for Genomics at the Flatiron Institute, Simons Foundation *"Decoding the human genome: from non coding variants to disease"*
6:05 p.m. – 6:25 p.m.	William Hwang, (Abstract Selected), Massachusetts General Hospital *"Multicellular spatial neighborhoods in pancreatic cancer are remodeled by neoadjuvant treatment"*
6:25 p.m. – 6:45 p.m.	Kevin King (Abstract Selected), University of California, San Diego *"Defining the ischemic borderzone with single cell spatial resolution"*
7:00 p.m. – 10:00 p.m.	**Welcome Reception** • Hilton Pool
Tuesday, June 7	**(Morning)**
7:30 a.m. – 9:00 a.m.	**Breakfast** • Waldorf Golf Pavilion
8:00 a.m. – 7:00 p.m.	**Meeting Registration / Hospitality Desk** • Bonnet Creek Ballroom Center Foyer
Plenary Session:	**Evolutionary Genomics (Stacey Gabriel, Broad Institute of MIT and Harvard, Chair)** • Floridian Ballroom Salon A-L
9:00 a.m. – 9:30 a.m.	Mary O'Connell, University of Nottingham, UK *"On the malleability of protein function"*
9:30 a.m. – 10:00 a.m.	Simone Immler, University of East Anglia, UK *"Consequences of haploid selection in sperm"*
10:00 a.m. – 10:30 a.m.	Elinor Karlsson, University of Massachusetts Medical School, and director of Vertebrate Genomics at the Broad Institute of MIT and Harvard *"The future of comparative genomics: finding meaning in sequence in a million genome age"*
10:30 a.m. – 11:00 a.m.	**Coffee Break** Sponsor Promenade
11:00 a.m. – 11:30 a.m.	Mike Snyder, Stanford University *"Transformative approaches to analyze genomes using machine learning reveals the genetic basis of complex diseases"*
11:30 a.m. – 11:50 p.m.	**Poster Flash Talks**
Tuesday, June 7	**(Afternoon)**
12:00 p.m. – 1:30 p.m.	**Gold Sponsor – Resolve Biosciences Workshop and Lunch** Jason T. Gammack, CEO and Co-founder of Resolve Biosciences Prof. Dr. Nikolaus Rajewsky Berlin Institute for Medical Systems Biology Michael Snyder, Ph.D. Department of Genetics, Stanford School of Medicine, Stanford University *"Context matters: Spatial- temporal subcellular gene expression patterns in systems biology and human cell atlases through spatial genomics."* • Bonnet Creek Ballroom Salon I-VI
12:00 p.m. – 1:30 p.m.	**AGBT Lunch** • Outdoor Waldorf Promenade
1:30 p.m. – 3:00 p.m.	**Poster Session with Coffee & Dessert** • Bonnet Creek Ballroom Salon VII-XII
Plenary Session:	**Genomics II (Martin Hirst, University of British Columbia, Chair)** • Floridian Ballroom Salon A-L
3:00 p.m. – 3:30 p.m.	Trevor Lawley, Microbiotica *"Making medicines from the human microbiome"*
3:30 p.m. – 4:00 p.m.	Alison Van Eenennaam, University of California, Davis *"Functional analysis and characterization of the hornless phenotype in cattle using genome editing"*
4:00 p.m. – 4:20 p.m.	Kiana Aran, (Abstract Selected), Cardea Bio *"From CPUs to BPUs: A new generation of transistors for rapid amplification-free genotyping"*
4:20 p.m. – 4:40 p.m.	Can Cenik, (Abstract Selected), University of Texas, Austin *"Single cell quantification of ribosome occupancy in early mouse development"*
Tuesday, June 7	**(Evening)**
5:00 p.m. – 7:15 p.m.	Dinner on your own
5:30 p.m. – 7:15 p.m.	**Women's Networking Event** • Signature Island

Concurrent Session:	**Cancer (Kim Doheny, Johns Hopkins University School of Medicine, Chair)** • Bonnet Creek Ballroom Salon I-VI
7:30 p.m. – 7:50 p.m.	Lukas Chavez, University of California, San Diego *"The landscape of extrachromosomal circular DNA in medulloblastoma"*
7:50 p.m. – 8:10 p.m.	Quan Nguyen, Institute for Molecular Bioscience, The University of Queensland *"Building a spatial single-cell multi-omics atlas and cellular interactome for skin cancer"*
8:10 p.m. – 8:30 p.m.	Michael Quail, Wellcome Sanger Institute *"NanoSeq: A high-throughput pipeline for the detection of somatic mutations at single-molecule resolution"*
8:30 p.m. – 8:50 p.m.	Billy Lau, Stanford University School of Medicine *"Massively parallel nanopore sequencing of circulating tumor DNA enables robust quantification of tumor burden"*
8:50 p.m. – 9:10 p.m.	Edwin Cuppen, Hartwig Medical Foundation *"Routine whole genome sequencing-based cancer diagnostics for precision medicine"*
9:10 p.m. – 9:30 p.m.	Panagiotis Karras, VIB Center for Cancer Biology *"Dissecting the melanoma ecosystem at single-cell resolution using Molecular CartographyTM"*

Concurrent Session:	**Computational Biology (Mike Zody, New York Genome Center, Chair)** • Floridian Ballroom Salon A-L
7:30 p.m. – 7:50 p.m.	Kristian Cibulskis, The Broad Institute *"Optimizing large-scale joint callsets for All of Us"*
7:50 p.m. – 8:10 p.m.	Melanie Kirsche, Johns Hopkins University *"Jasmine: Population-scale structural variant comparison and analysis"*
8:10 p.m. – 8:30 p.m.	Vinicius Medeiros Fava, McGill University *"A system biology approach identifies candidate drugs to reduce mortality in severely ill COVID-19 patients"*
8:30 p.m. – 8:50 p.m.	Tanner Jensen, Stanford University *"Integrative analysis of long-read and functional genomics data improves prioritization of pathogenic variants in exome-negative, rare-disease patients"*
8:50 p.m. – 9:10 p.m.	Jialong Jiang, California Institute of Technology *"Spin-network models identify and predict human immunomodulatory drug response signatures in high-throughput single-cell mRNA-seq data"*
9:10 p.m. – 9:30 p.m.	Chandrima Bhattacharya, Weill Cornell Medicine *"Microbial survival in the Anthropocene: Bioremediation and biomining potential of a superfund site - the Gowanus Canal"*

9:30 p.m.	**CLICK2WIN (Formerly Passport to Prizes)** Sponsor Promenade
Wednesday, June 8	**(Morning)**
7:30 a.m. – 9:00 a.m.	**Breakfast** • Waldorf Golf Pavilion
8:00 a.m. – 7:00 p.m.	**Hospitality Desk** • Bonnet Creek Ballroom Center Foyer
Plenary Session:	**Genetics I (Federica Di Palma, Genome British Columbia, Chair)** • Floridian Ballroom Salon A-L
9:00 a.m. – 9:30 a.m.	Martin Hirst, University of British Columbia *"SMARCB1 loss is associated with compartment switching and sensitivity to NSD1/2 inhibition"*
9:30 a.m. – 10:00 a.m.	Zach Lippman, Cold Spring Harbor Laboratory (CSHL) and Howard Hughes Medical Institute (HHMI) *"Dynamic diversification of gene duplications in shaping trait variation"*
10:00 a.m. – 10:20 a.m.	Lesley Shirley, (Abstract Selected), University Wellcome Sanger Institute *"Large-scale whole-genome sequencing of SARS-CoV-2 at the Sanger Institute: The COVID-19 Genomics UK consortium"*
10:20 a.m. – 11:00 a.m.	**Coffee Break** • Sponsor Promenade
11:00 a.m. – 11:20 a.m.	Samantha Sholes (Abstract Selected), Johns Hopkins University School of Medicine *"Nanopore sequencing telomere length measurement method reveals chromosome specific telomere lengths"*
11:20 a.m. – 11:40 a.m.	**Poster Flash Talks**
Wednesday, June 8	**(Afternoon)**
12:00 p.m. – 2:00 p.m.	**Silver Sponsor Workshops and Lunch** • Bonnet Creek Ballroom Salon I-VI
12:00 p.m. – 1:00 p.m.	Silver 1 Sponsor Workshop - Ultima Genomics Gilad Almogy, CEO and Co-Founder, Ultima Genomics *"Welcome and introduction"*

Doron Lipson, CSO, Ultima Genomics
"Enabling genomic analysis at scale: applications and collaborations"

Stacey Gabriel, Broad Institute
"An evaluation of the Ultima Genomics sequencing platform: high-throughput sequencing at low-cost"

Reuben Saunders, Whitehead Institute
"Mapping the genetic landscape of complex cellular phenotypes with genome-scale Perturb-Seq"

Erez Lieberman Aiden, Baylor College of Medicine
"Generating 3D maps of the human genome at basepair resolution using the Ultima Genomics platform reveals principles of enhancer-promoter wiring"

1:05 p.m. – 2:05 p.m. **Silver 2 Sponsor Workshop and Lunch – Element Biosciences**
Moderated by Shawn Levy
Andrew Carroll (Deep Variant), Chris Mason (Weil Cornell), Sheila Dodge (Broad), and Mike Previte (Element Biosciences)
"Elements of freedom: astronauts, forensics, and state of the art genomics"

12:00 p.m. – 2:00 p.m. **AGBT Lunch**
• Outdoor Waldorf Promenade

Bronze Sponsor Workshops
• Floridian Ballroom Salon A-L

(Stephen Montgomery, Stanford University School of Medicine, Chair)

2:15 p.m. – 2:35 p.m. **Bronze 1 Sponsor – Twist Bioscience**
Emily Leproust, PhD
"Continuing the relentless pursuit of precision medicine - new tools for cancer genomics research"

2:35 p.m. – 2:55 p.m. **Bronze 2 Sponsor – PacBio**
David Miller, Vice President, Product Marketing, PacBio
"Sequencing by Binding (SBB®) delivers unprecedented NGS accuracy"

2:55 p.m. – 3:15 p.m. **Bronze 3 Sponsor – DNA Script**
Thomas Ybert and Xavier Godron
"Same-Day enzymatic DNA synthesis: now print primers and labeled probes using the SYNTAX System"

3:15 p.m. – 3:30 p.m. **Bronze 4 Sponsor – Integrated DNA Technologies**
Dr. Steven Henck, Vice President of R&D, IDT
"Introducing IDT's more complete NGS portfolio. xGen™ NGS – made for you."

3:30 p.m. – 3:45 p.m. **Bronze 5 Sponsor – Miroculus**
Fay Christodoulou
"Miro Canvas: an automation solution for every lab"

3:45 p.m. – 4:00 p.m. **Bronze 6 Sponsor – Agilent**
Bellal Moghis, Director of Marketing
"Paving the way forward | Leveraging Agilent's innovations to achieve new heights"

4:00 p.m. – 4:15 p.m. **Bronze 7 Sponsor – New England Biolabs**
Bradley W. Langhorst, Ph.D., Product Development Group Leader
"Clearing the CoV Fog: Pandemic LAMP and Sequencing at NEB"

4:15 p.m. – 4:27 p.m. **Bronze 8 Sponsor – Bruker**
Mark R. Munch, Ph.D. - President Bruker NANO & CEO Acuity Spatial Genomics, Inc.
"Bruker presents novel cutting-edge solutions for deeper, quantitative analysis of spatial biology and single cell omics"

4:27 p.m. – 4:39 p.m. **Bronze 9 Sponsor – Bionano Genomics**
Alka Chaubey, Ph.D., Chief Medical Officer, Bionano
"OGM + NGS for transforming genome biology. The FUTURE is Bright!"

4:40 p.m. – 6:10 p.m. **Poster Session and Wine Reception**
• Bonnet Creek Ballroom Salon VII-XII

Wednesday, June 8

(Evening)

6:15 p.m. – 7:25 p.m. **AGBT Dinner**
• Outdoor Waldorf Promenade

Concurrent Session: **Technology (Chris Mason, Weill Cornell Medicine, Chair)**
• Floridian Ballroom Salon A-L

7:30 p.m. – 7:50 p.m. Chongyuan Luo, University of California, Los Angeles
"In situ single-cell epigenomic profiling by photonic-indexing sequencing"

7:50 p.m. – 8:10 p.m. Sheridan Cavalier, Johns Hopkins School of Medicine Dept of Neuroscience
"Single-cell transcript isoform sequencing of the actived adult mouse hippocampus with 10x Genomics and Oxford Nanopore"

8:10 p.m. – 8:30 p.m. George Emanuel, Vizgen
"Molecular atlassing with MERSCOPE: A new spatial transcriptomics technique accurately reveals the organization of the mouse brain transcriptome"

8:30 p.m. – 8:50 p.m. Heonseok Kim, Stanford University
"A single cell CRISPR and nanopore long read approach for modifying mRNA sequence and generating novel transcript isoforms"

8:50 p.m. – 9:10 p.m. Joseph Beechem, NanoString Technologies Inc.
"High-plex (> 1000), multi-omic (RNA plus protein), spatial molecular imaging with sub-cellular resolution in FFPE tissue"

9:10 p.m. – 9:30 p.m. Christopher Grochowski, Baylor College of Medicine
"Resolving complex genomic inversion structures"

Concurrent Session: **Biology (Penelope Bonnen, Baylor College of Medicine, Chair)**
• Bonnet Creek Ballroom Salon I-VI

7:30 p.m. – 7:50 p.m. Jeffrey Shaman, Coriell Life Sciences
"Results and reflections of a real-world implementation of a pharmacogenomics-empowered comprehensive medication management (CMM) program"

7:50 p.m. – 8:10 p.m. Pengfei Liu, Baylor College of Medicine
"Analytical evaluation of clinical whole genome sequencing with a highly novel, vastly cheaper DNA sequencing-by-synthesis technology"

8:10 p.m. – 8:30 p.m. Varuna Chander, Baylor College of Medicine
"Somatic mosaicism and cardiovascular disease: Uncovering hidden genetic etiology with clinical Implications"

8:30 p.m. – 8:50 p.m. Eloi Schmauch, Massachusetts Institute of Technology, University of Eastern Finland, MIT and Harvard, Broad Institute
"Single-cell spatio-transcriptional dissection of human heart in health and coronary artery disease reveals cell-type-specific driver genes and pathways"

8:50 p.m. – 9:10 p.m. Alma Andersson, Science For Life Laboratory (SciLifeLab) and Royal Institute of Technology (KTH)
"Building common coordinate frameworks for spatial transcriptomics data to promote modelling of large and diverse datasets"

9:10 p.m. – 9:30 p.m. Hilde Nelissen, Ghent University, Department of Plant Biotechnology and Bioinformatics / VIB Center for Plant Systems Biology
"Molecular CartographyTM reveals novel cell types and gene functions in the maize shoot apex"

Thursday, June 9 (Morning)

7:30 a.m. – 9:00 a.m. **Breakfast**
• Waldorf Golf Pavilion

8:00 a.m. – 6:00 p.m. **Hospitality Desk**
• Bonnet Creek Ballroom Center Foyer

Plenary Session: 1:30pm **Genetics II (Len Pennacchio, Lawrence Berkeley National Laboratory and DOE Joint Genome Institute, Chair)**
• Floridian Ballroom Salon A-L

9:00 a.m. – 9:30 a.m. Neil Hanchard, National Human Genome Research Institute
"MIA: uncaptured sequence variation in african populations"

9:30 a.m. – 10:00 a.m. Alicia Zhou, Chief Science Officer, Color
"Deploying genomics at population scale"

10:00 a.m. – 10:20 a.m. Federica De Palma, (Abstract Selected) Genome British Columbia, Vancouver CA / University of East Anglia, UK
"Evolution of regulatory networks associated with traits under selection in cichlids"

10:20 a.m. – 11:00 a.m. **Coffee Break** Sponsor Promenade

11:00 a.m. – 11:20 a.m. Joshua Cohen, (Abstract Selected), Johns Hopkins University School of Medicine
"Detection of low-frequency DNA variants by targeted sequencing of the Watson and Crick strands"

11:20 a.m. – 11:40 a.m. Stephanie Yan, (Abstract Selected), Johns Hopkins University
"A complete reference genome improves analysis of human genetic variation"

12:00 p.m. – 1:00 p.m. **Silver 3 Sponsor Workshop and Lunch – 10 X Genomics**
Peter Smibert, Mike Gibbons, Augusto Tentori and Francesca Meschi
"The next generation of single cell and spatial technologies"

• Bonnet Creek Ballroom Salon I-VI

12:00 p.m. – 1:30 p.m. **AGBT Lunch**
• Outdoor Waldorf Promenade

Thursday, June 9 **(Afternoon)**

Plenary Session: **Technology (Penelope Bonnen, Baylor College of Medicine, Chair)**
• Floridian Ballroom Salon A-L

1:30 p.m. – 2:00 p.m. Fei Chen, The Broad Institute of MIT and Harvard
"Tissue genomics: placing genomic measurements in context"

2:00 p.m. – 2:30 p.m. Steven Reilly, Yale School of Medicine
"Towards understanding the impact of all variants: identifying 3'UTR causal variants for human disease and evolution"

2:30 p.m. – 2:50 p.m. Gilad Almogy, Ultima Genomics
"Novel large-scale sequencing platform using mostly-natural sequencing by synthesis"

2:50 p.m. – 3:20 p.m. **Coffee Break**
• Floridian Ballroom Foyer

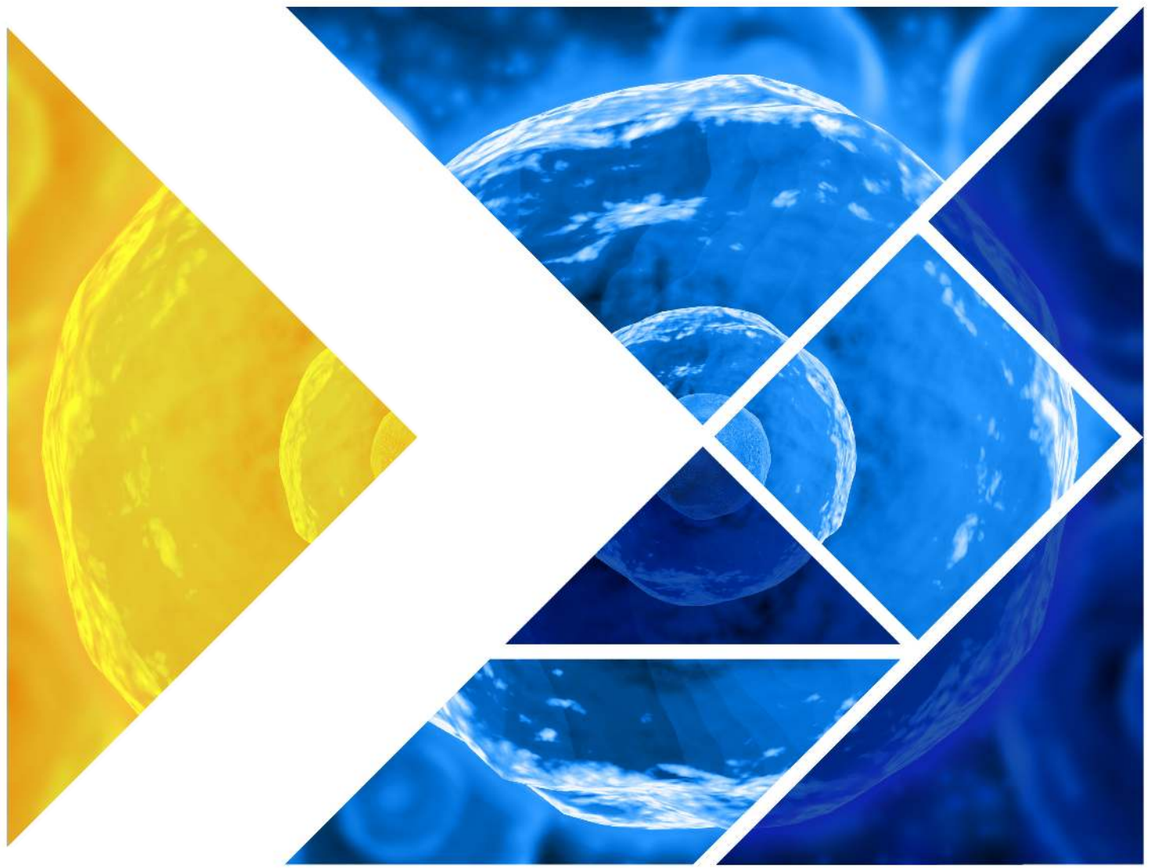
3:20 p.m. – 3:40 p.m. Ghamdan Al-Eryani, (Abstract Selected), Garvan Institute of Medical Research
"Multi-omic integration of single-cell and spatial RNA and protein analysis identifies novel tumor-infiltrating lymphocyte phenotypes"

3:40 p.m. – 4:00 p.m. Shruti Iyer, (Abstract Selected), Cold Spring Harbor Laboratory/Stony Brook University
"ACME: An Affinity-based Cas9 Mediated Enrichment method for targeted nanopore sequencing"

4:00 p.m. – 4:20 p.m. Christopher Vollmers (Abstract Selected), University of California, Santa Cruz
"Illumina but with Nanopore"

4:20 p.m. – 4:30 p.m. **Click2Win Winner Announced**
Closing Comments and Meeting Feedback

6:00 p.m. – 10:00 p.m. **Farewell Dinner Party**-Luau!
• Waldorf Golf Pavilion



Abstract Table of Contents

AGBT2022™
GENERAL MEETING

Cancer Omics

Multi-omics modeling of injury-driven microenvironment trajectory in liver cancer

Jiwoon Park[1,2], **Charles M. Rice**[2], **Christopher E. Mason**[1,3,4]

1 Department of Physiology, Biophysics and Systems Biology, Weill Cornell Medicine, New York, NY
2 Laboratory of Virology and Infectious Disease, the Rockefeller University, New York, NY
3 Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY
4 The Feil Family Brain and Mind Research Institute, Weill Cornell Medicine, New York, NY

Carcinogenesis often involves a systemic process where the delicate regulation between metabolism, tissue structure, and the immune system is disrupted. For example, upon injury from viruses or toxic materials, the liver activates immune system and initiates regenerative processes. Yet, if the injury persists, the accumulated inflammation and aberrant proliferation disrupts the tissue architecture and cellular microenvironment, causing advanced disease states (i.e., fibrosis and cirrhosis in liver) and cancer in many cases. Although distinct genotoxic insults and etiologies are defined, how the process differs by injury has not yet been characterized. With large patient population of liver cancer in mind (42,230 patients newly diagnosed with liver cancer with 5-year relative survival of 20.3%), here we present novel computational tools and spatial omics techniques to reveal the drivers of these tissue disruptions. We find that different injuries – viral hepatitis or fatty liver diseases – create their own unique trajectories of immune system activation and dysfunction, disease progression, and subtle impacts on the tissue and subsequent disease states. By analyzing liver biopsy transcriptomic data obtained from healthy, hepatitis C virus infected, nonalcoholic fatty liver, and liver cancer patients, we define immune cell types and pathways specific to each disease etiologies and states. We also find that CTNNB1 and TP53, two of the top mutated genes in liver cancer, mutates exclusively in viral and fatty liver associated cancer samples respectively. We furthermore integrate all the transcriptome datasets in different states and model with graph neural network, where expression levels as features and connectivity of the samples as edges, and perform disease trajectory classifications. We believe that the results and analytical techniques acquired from this system can be generalized to other organs such as kidney, lung, or heart and to understand tumor immune microenvironment and developing appropriate treatment strategies for cancer patients.

Assessment of the TNF-a pathway using Molecular CartographyTM analysis on cytoreductive radical prostatectomy samples

Cancer Omics

POSTER 101

Isaac Kim[1], **Nachiket Kashikar**[4], **Antonina Mitrofanova**[2], **Jeffrey Rosenfeld**[3,*]

1 Yale School of Medicine, New Haven, CT
2 Department of Biomedical and Health Informatics at Rutgers School of Health Professions, New Brunswick, NJ
3 Department of Pathology and Lab Medicine, Rutgers University, New Brunswick, NJ
4 Resolve Biosciences, Monheim am Rhein, Germany
*Presenting author

Although prostate cancer is largely curable, metastatic prostate cancer remains lethal and comprise approximately 6% of new prostate cancer diagnoses each year. For men with de novo metastatic prostate cancer, systemic therapy based on androgen deprivation is the standard of care and surgery is not part of the accepted treatment regimen. Notwithstanding, there is a body of literature that supports local surgery, cytoreductive radical prostatectomy, in men with metastatic prostate cancer. To start exploring the role of cytoreductive radical prostatectomy, we launched a phase 1 safety and feasibility study in 2015. The study confirmed safety as well as heterogeneous oncologic response where some had spectacular response while others experience rapid disease progression. To start developing a rational strategy for personalizing cytoreductive radical prostatectomy, we carried out a bulk RNAseq analysis and found an association between the TNF-a pathway and treatment outcome. To further assess the underlying mechanism, we performed 3D localization of transcriptomics (Molecular Cartography™, Resolve Biosciences) in radical prostatectomy specimens obtained via the clinical trial. The results confirmed the feasibility of such approach that will likely add to our armamentarium in interrogating the biology of prostate cancer.

Spatial characterization of the tumor microenvironment in pancreatic cancer with whole transcriptome profiling and high-plex single-cell imaging

Cancer Omics

POSTER 102

Youngmi Kim[1], **David T. Ting**[2], **Amaya Pankaj**[2], **Eunae You**[2], **Sean Kim**[1], **Mark Gregory**[1], **Nathan Schurman**[1], **Vikram Devgan**[1], **Sarah Warren**[1]

1 NanoString Technologies, Seattle, WA
2 Massachusetts General Hospital, Boston, MA

Repeat sequences have been shown to play a role in wide range of diseases including cancer, chronic viral infection, and neurodegeneration. A recent Phase II clinical trial of pancreatic cancer indicating that expression of certain repeat RNAs are correlated with response to combined immune checkpoint blockade and sublethal radiation therapy. More mechanistic studies have now revealed that these repeat elements are transcribed and often packaged in extracellular vesicles (EVs) which can be released from the tumor cells and interact with cells in the tumor microenvironment including cancer-associated fibroblasts (CAFs). Thus, it is important to understand the spatial distribution of repeat RNAs found in tumor cells and the surrounding CAFs that can provide mechanistic insight into their consequence in shaping the cellular response to tumors. Of particular interest is the impact they have on CAF subtypes that have been defined by single cell RNA-seq (scRNA-seq), which include the inflammatory (iCAF) or myofibroblastic (myCAF) phenotypes. Given the limitations of bulk RNA-seq and scRNA-seq in understanding the precise spatial distribution of this phenomenon, we used the NanoString GeoMx and Spatial Molecular Imager (SMI) to visualize these repeat elements and study their impact on the surrounding cells.

In this study, we used the GeoMx Whole Transcriptome Atlas and RNAscope to visualize repeat elements and characterize the impact of repeat elements-presenting regions of pancreatic cancer. Then, we used SMI to collect gene expressions of spatially-annotated single cells in both repeat elements-positive and repeat elements-negative regions. SMI uses fluorescent molecular barcodes to enable in-situ measurement of biological targets on an intact tissue sample, such as formalin-fixed paraffin-embedded tissues using a 1000+-plex RNA panel. Using the RNA expression data from SMI, we were able to spatially visualize how EVs encapsulating repeat elements impact the surrounding CAFs and other cells in the microenvironment. Using cell type deconvolution algorithms, we characterize the changes in gene expression within cell populations using the GeoMx as well as individual cells using SMI. Using nearest neighbor analyses, we are able to characterize expression profile changes between CAFs and immune cells that are close and distant from repeat expression tumor cells.

These results have provided a new lens in viewing the impact of repeat element tumor expression and mechanistic role in the microenvironment via EVs, which have the opportunity to identify novel therapeutic avenues and biomarkers in pancreatic cancer.

For research use only, not for use in diagnostic procedures.

Cancer Omics

POSTER 103

Breaking the 100-plex protein spatial single-cell imaging barrier in FFPE tissues

Zachary Lewis[1], **Tiên Phan-Everson**[1], **Gary Geiss**[1], **Mithra Korukonda**[1], **Ruchir Bhatt**[1], **Carl Brown**[1], **Dwayne Dunaway**[1], **Joseph Phan**[1], **Alyssa Rosenbloom**[1], **Brian Filanoski**[1], **Rhonda Meredith**[1], **Kan Chantranuvatana**[1], **Yan Liang**[1], **Emily Brown**[1], **Brian Birditt**[1], **Giang Ong**[1], **Hye Son Yi**[1], **Erin Piazza**[1], **Vikram Devgan**[1], **Patrick Danaher**[1], **Michael Rhodes**[1], **Joseph M. Beechem**[1]

1 NanoString® Technologies, Seattle WA

Spatial biology, together with a very new focus on high plex protein-detection technology, is the next frontier of molecular research, because it provides biological insight in the context of tissue, cellular location, subcellular organization, and (ultimately) function. The CosMx™ Spatial Molecular Imager is the first platform to demonstrate simultaneous single-cell and sub-cellular detection of over 100 proteins on standard, biobanked, FFPE tissue samples. This high plex protein panel detects key drivers of cancer progression and immune cell activation states. Key to CosMx protein technology is an antibody-oligonucleotide-conjugate 64-bit encoding method, not a cyclic exchange method. This chemistry has a direct migration-path to 1000-plex, without any change in antibody-conjugates, cyclic chemistry reagents, or detection platform. The encoding scheme is enabled by a 20nm-sized, hybridization-based optical barcode. The CosMx system uses a fully automated, cyclic microfluidics imaging system, high-resolution optics and 3D capability. The raw cyclic encoded 4-color tissue images are decoded using an error-robust decoding algorithm that detects protein (sub)-cellular localization and quantifies expression level. CosMx oligo-labeled antibodies undergo rigorous QC testing, site-specific oligo labeling chemistry, and size-exclusion purification to select for pure antibody-oligo imaging reagent with no unconjugated antibody or free oligonucleotide contamination. The CosMx protein assay reagents were validated on multi-organ FFPE tissue microarrays and 45 human FFPE cell lines, including overexpression lines for key targets and cellular activation states, such as GITR, CD278, PD-L1, and PD-1. Benchmarking to multiple orthogonal datasets (e.g., the Human Protein Atlas, Cancer Cell Line Encyclopedia, low-plex IHC) demonstrating that the assay is highly sensitive and specific. CosMx SMI produces protein localization maps for each target that can be integrated with SMI's (multi-omic) 1000-plex RNA-detection, generating an unprecedented information-rich view of spatial biology that could usher-in a completely new understanding of health and disease.

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Cancer Omics

POSTER 104

Optimizing promoter sequencing in cancer genomes

Nicole Francis[1], **Meifang Qi1**, **Katie Larkin**[1], **Gina Vicente**[1], **Matt DeFelice**[1], **Stacey Gabriel**[1], **Niall Lennon**[1], **Carrie Cibulskis**[1], **Esther Rheinbay**[1,2]

1 Broad Institute of MIT and Harvard, Cambridge, MA
2 Massachusetts General Hospital Cancer Center, Charlestown, MA

Gene panel and Exome sequencing are commonly used techniques in Cancer genomics and have been applied to hundreds of thousands of tumor specimens in the clinical and research settings. While these techniques are highly advanced and optimized for the coding regions of genes, there currently exists no platform to systematically evaluate promoter regions. Whole-genome sequencing, which in principle allows investigation of promoter driver events, is still costly and suffers from systematic under-sampling of GC-rich promoter regions. Whole genomes will also likely continue to fall short of the coverage goals required for detection of low frequency events such as in liquid biopsy specimens. As a result of the technology limitations described above, relatively few recurrent driver mutations in promoters have to date been described (TERT, FOXA1, indels in the TP53 5'UTR/promoter).

To alleviate these limitations and allow for a systematic discovery, we have undertaken an effort to design and optimize a targeted capture assay specifically for functional promoter regions (n=3203) of cancer genes (n=2363) along with a computational workflow to analyze and characterize somatic variations. Recent improvements in targeting methodology with TWIST have generated more even coverage across the spectrum of GC-content allowing for a view into these previously 'dark regions' of the cancer genome. Optimization of library construction and hybridization conditions for performance of this challenging target yielded conditions with a reasonable balance in coverage of high priority targets while still allowing for efficient, cost effective capture conditions (> 85% of reads on target). We will describe our technical methods and results from the assay optimization as well as initial cell line sequencing analysis and promoter region variant discovery.

M2GEN's oncology-focused analytics accelerate the discovery, development, and delivery of cancer research using a custom xGenTM whole exome research panel and gene expression profiling

Cancer Omics

POSTER 105

Zhijie Jiang[1], **Qi Zhang**[1], **McKayla J. Riggs**[2], **Jill M. Kolesar**[3], **Dan S. Hughes**[1], **Oliver A. Hampton**[1],

1 Department of Bioinformatics and Biostatistics, M2GEN, Tampa, FL
2 Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, University of Kentucky, Lexington, KY
3 Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, College of Pharmacy, Markey Cancer Center, University of Kentucky, Lexington, KY

Whole exome sequencing (WES), targeted-panel sequencing, gene expression profiling (RNAseq) and the companion analytic pipelines are widely accepted industry standards as powerful methods to discover potential causative variants that drive diseases such as cancer. WES and RNAseq's clinical research applications include evaluating genetic sequences for therapeutic drug targeting or identifying biomarkers for clinical trial stratification and/or personalized treatment. Oncology research leverages both WES and RNAseq of patients' tumors to advance progress toward improved cancer diagnoses, prognoses, and personalized treatment.

M2GEN's custom xGenTM whole exome panel includes a custom probe spike-in that targets select oncology-actionable genes such as promoter, enhancer, and intronic regions for structural variant detection. M2GEN's companion analytic pipelines invoke industry standard algorithms from the public and private sectors, including GATK and Sentieon®, and leverage the ORIEN AVATAR® dataset of tens of thousands match tumor and germline molecular and clinical data for enhanced calling of somatic mutations and gene expression profiles.

In this use-case of M2GEN's endometrial cancer cohort of 987 patients as of July 2021, two separate findings are presented: (1) DACH1 mutation frequency in endometrial cancer is associated with high tumor mutation burden; and (2) mutational signatures in endometrial cancer reveal expanded proof-reading deficiency within POLE and POLD1. In both finding, M2GEN's enhanced analytics provide improved somatic mutation assessment of cancer patients' tumors, aiding M2GEN's mission to accelerate the discovery, development, and delivery of more personalized therapies.

Comprehensive single-cell spatial atlas of NSCLC: public release of 1 million cell data set generated on Spatial Molecular Imager

Cancer Omics

POSTER 106

Joseph Beechem, **Youngmi Kim, Patrick Danaher, Sean Kim, Sarah Murphy, Nathan Schurman, Michael Leon, Emily Killingbeck, Bryce Kan, Tami Gilderman-Ramirez, Michael Rhodes**

NanoString® Technologies, Seattle, WA

Understanding the multi-cellular spatial interactions that create organs and organisms (in both health and disease) is one of the key challenges of modern biology. Advanced sequencing technologies have allowed the analysis of large numbers of dissociated cells, enabling projects to analyze all the cell types in individual organs. However, dissociated cells lose the spatial information on cellular organization and interaction. The spatial arrangement of cells and their local spatial environments is increasingly recognized as being essential to understanding biological function in both health and disease. The non-small cell lung cancer (NSCLC) atlas will measure more than 1000 different RNA targets simultaneously in-situ, with subcellular resolution and spatial mapping in 3 dimensions. The CosMx SMI system was used to analyze FFPE samples from 5 NSCLC patients, generating data on over 1 million cells. To allow the entire scientific community to explore these data, and develop new analysis methodologies and algorithms, the data has been made available to all researchers. To our knowledge, this is the largest biobanked sample high-plex RNA data set ever collected. This data is now available to all researchers to allow the entire scientific community to develop new analysis methodologies and discover new insights into NSCLC.

Our initial analysis of this data reveals tumor biology in unprecedented sub-cellular and single-cell spatial detail. Using this dataset, we have observed the complex ways in which 18 immune cell populations invade tumors: in one tumor, memory CD8 T-cells invade broadly throughout the tumor, while neutrophils gather densely in large stroma areas. The data enable the delineation of niches within the tumor microenvironment, such as lymphoid structures packed with B-cells and T-cells, pockets of dense macrophages, stroma dominated by two subpopulations of fibroblasts, and tumor interior with only sparsely invading immune cells. We describe how immune cells change behavior across spatial contexts, contrasting expression of macrophages invading diffusely into the tumor and macrophages gathering densely in the stroma. Furthermore, we characterize the interaction of tumor cells and the environment, e.g. one tumor up-regulates VEGFA and NRDG1 only when contacting stroma.
The data is available at www.nanostring.com

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Clinical whole exome and companion RNA sequencing enables expansion of personalized cancer vaccine clinical trials

Cancer Omics

POSTER 107

Junko Tsuji[1], **Alyssa Macbeth**[1], **Mark Fleharty**[1], **Caroline Petersen**[1], **Christopher Kachulis**[1], **Takuto Sato**[1], **Micah Rickles-Young**[1], **Caroline Petersen**[1], **Maegen Harden**[1], **Katie Larkin**[1], **Gina Vicente**[1], **Nicole Francis**[1], **Amitha Sandur**[1], **Catherine J. Wu**[1,2,3,4], **Patrick Ott**[2,3,4], **Maegan Harden**[1], **Chip Stewart**[1], **Carrie Cibulskis**[1], **Gad Getz**[1,4,5], **Niall Lennon**[1], **Stacey Gabriel**[1]

1 Broad Institute of MIT and Harvard, Cambridge, MA
2 Dana-Farber Cancer Institute, Boston, MA
3 Brigham and Women's Hospital, Boston, MA
4 Harvard Medical School, Boston, MA
5 Massachusetts General Hospital, Boston, MA

There is great promise in treating cancer patients with personalized vaccines that are designed to stimulate a highly specific anti-tumor immune response. Since 2016, phase I clinical trials designed to test vaccines targeting personal tumor neoantigens (hereafter, NeoVax), manufactured in real-time through identification of peptides derived from tumor-specific mutations predicted to be immunogenic, have been completed and shown to be highly feasible and immunogenic in patients with melanoma. Over the past few years, our portfolio of trials have rapidly expanded to other tumor types,including renal cell carcinoma, glioblastoma, and ovarian carcinoma in addition to melanoma. Support of such trials requires robust, high quality sequencing and complex analytical pipelines with accurate somatic mutation calling and neoantigen prediction.

In this presentation, we demonstrate the improved data quality of our bundled clinical whole-exome sequencing (WES) with TWIST targeting strategy coupled with unique molecular indexed TWIST exome enriched RNA sequencing and the performance of an optimized neoantigen prediction pipeline. Our assay, with the introduction of the new target enrichment protocol, showed stable coverage across targeted genomic regions including the regions that are difficult to capture due to low or high GC content. The coverage uniformity was observed in various specimen types such as fresh frozen tissues, formalin-fixed paraffin-embedded specimens, and liquid biopsy with cell-free tumor DNA. To obtain the best performance of mutation calls and neoantigen predictions on the new assay, we have benchmarked and optimized our Dragen variant detection and cloud-based analytical pipeline components for supporting the next generation of expanded NeoVax trials. Dragen reached 92.7% and 86.8% SNV and INDEL detection sensitivity at the simulated allelic fraction ⩾ 5% at 175X mean target coverage (95.2% for SNV with ⩾ 10% allelic fraction and 93.9% for INDEL with ⩾ 20% allelic fraction at sites covered at least 175X) on our benchmarking data created from mixed HapMap samples.

Genomic profiling of canine cancers by tumor and liquid biopsy analysis for comparative oncology

Cancer Omics

POSTER 108

Ilya Chorny, **Kristina M. Kruglyak, John A. Tynan, Prachi Nakashe, Gilberto E. Hernandez, Rita Motalli-Pepio, Lisa McLennan, Lauren E. Holtvoigt, Jason Chibuk, Andi Flory, Daniel S. Grosu, Dana W.Y. Tsui**

PetDx, Inc., La Jolla, CA

Characterizing actionable variants in humans and dogs can expand the knowledge base of comparative oncology and support the development of novel treatments for the benefit of both species. Genomic profiling via tumor biopsy confers risk to the patient in both humans and dogs and is often complicated by tumor heterogeneity. Blood-based liquid biopsy has been shown to open new opportunities to profile the cancer genome noninvasively in both species.1,2

In this study, human and canine cancer genes from the Cancer Gene Census (CGC) were evaluated for homology, including oncogenes and tumor suppressor genes taken from COSMIC v913. Cancer-associated genes were found to have higher homology between humans and dogs (> 90%) than the remainder of the genes in the genome (< 85%). The top 100 human DNA single nucleotide variants from the COSMIC database1 that had a canine orthologue were then studied in a cohort of 350 dogs that had a variety of cancers including lymphoma, soft-tissue sarcoma, hemangiosarcoma, osteosarcoma and others. Most patients had multiple tissue samples collected from spatially separated sites in the tumor to study the presence of tumor heterogeneity. Corresponding blood plasma samples, collected from each patient prior to the surgical resection of the tumor, were sequenced and the genomic profiles of the blood and tumor samples were compared. A large number of actionable human cancer variants4 were found to be present in dogs and were noninvasively detected using liquid biopsy and in corresponding tumor samples. This research demonstrates the potential for translation of knowledge between humans and dogs to support the development of novel treatments to benefit both species.

References:
1. Chibuk J, Flory A, Kruglyak KM, Leibman N, Nahama A, Dharajiya N, et al. Horizons in Veterinary Precision Oncology: Fundamentals of Cancer Genomics and Applications of Liquid Biopsy for the Detection, Characterization, and Management of Cancer in Dogs. Frontiers Vet Sci. 2021;8:664718.

2. Kruglyak KM, Chibuk J, McLennan L, Nakashe P, Hernandez GE, Motalli-Pepio R, et al. Blood-Based Liquid Biopsy for Comprehensive Cancer Genomic Profiling Using Next-Generation Sequencing: An Emerging Paradigm for Non-invasive Cancer Detection and Management in Dogs. Frontiers Vet Sci. 2021;8:704835.

3. COSMIC - Catalogue of Somatic Mutations in Cancer [Internet]. Cancer.sanger.ac.uk. 2021 [cited 12 July 2021]. Available from: https://cancer.sanger.ac.uk/cosmic

4. OncoKB [Internet]. Oncokb.org. 2021 [cited 12 July 2021]. Available from: https://www.oncokb.org/

The ResolveOME platform: integrated whole genome and whole transcriptome profiling from a single cell to unlock drug resistance mechanisms

Cancer Omics

POSTER 109

Tatiana V. Morozova, **Jeff G. Blackinton, Isai Salas-González, Viren R. Amin, Victor J. Weigman, Jon S. Zawistowski, Gary L. Harton, Jay A.A. West**

BioSkryb Genomics, Inc., Durham, NC

To elucidate drug resistance mechanisms in cancer and to deconvolve the molecular basis of variability in treatment responses, it is paramount to apply approaches that integrate multi-omics data. Existing methodologies aiming to couple genomic and transcriptomic information are limited in that the genomic information is based on a targeted approach, providing insight into only a small fraction of the genome. We have developed a novel, true multi-omic platform, ResolveOME, that combines whole genome amplification with transcriptome profiling from the same single cell. The platform unifies template-switching single-cell RNAseq chemistry with modified ResolveDNA™ whole genome amplification (WGA) technology based on Primary Template-directed Amplification (PTA). The workflow then relies on affinity purification of first-strand cDNA and subsequent separation of the RNA/DNA fractions to allow for independent library preparation of the fractions followed by ResolveDNA™ library preparation and sequencing. To demonstrate the validity of this platform we generated GM12878 cell data using the ResolveOME chemistry and compared it to data from the ResolveDNA™ WGA kit and to bulk RNA sequencing. Products from the DNA and RNA arms of the protocol demonstrated comparable product sizes and consistent yields to the bulk RNAseq and standard ResolveDNA™ chemistry. Low-pass sequencing of the DNA arm of the ResolveOME workflow showed robust performance, including high genomic coverage and high library diversity. Gene expression analysis of the RNA arm of ResolveOME revealed the ability to detect ~ 10K expressed genes in bulk RNA samples and ~ 8K expressed genes in the single cells, concordant with published data for GM12878 cells. Our next focus is to extend the ResolveOME platform to established cancer drug resistant models in cell lines to identify SNVs, differentially expressed transcripts, and CNV contributing to drug resistance, as well as to primary patient samples. We have demonstrated here that the ResolveOME platform provides unprecedented, and non-targeted multi-omics data from individual single cells. The utility of ResolveOME extends to all applications which require both whole genome and transcriptome data from a single cell and/or samples with limited biological material.

Precise high-plex spatial immune cell profiling with ChipCytometry™

Cancer Omics

POSTER 110

Thomas Campbell, PhD*

Canopy Biosciences, St. Louis, MO
*Presenting Author

Immunohistochemistry (IHC) is the most widely used diagnostic technique in tissue pathology. However, standard IHC is associated with several limitations including the labeling of a single marker per tissue section and limited quantification of cell populations. The limited multiplexing results in missed opportunities to ask key questions about tumor biology that could be important for diagnostic and prognostic outcome. ChipCytometry™ is a novel precise spatial multiplexing imaging platform that addresses these challenges by combining iterative immuno-fluorescent staining with high-dynamic range imaging to facilitate quantitative phenotyping with single-cell resolution. The platform enables simultaneous detection of multiple markers on a single tissue section and enable accurate quantification of protein expression levels are necessary to deeply profile single cells, understand interactions between key immune cells, and identify topographic biomarkers. Standard FCS files are generated from multichannel OME-TIFF images, enabling identification of cellular phenotypes via flow cytometry-like hierarchical gating. Quantification of results reveal precise expression levels for each marker in the assay in each individual cell in the sample, while maintaining spatial information about each cell. ChipCytometry™ has the potential to inform the discovery of novel biomarkers for immuno-oncology research and is likely to have a significant impact on precision medicine and human health in the future.

Presenter Bio: Thomas Campbell is a Product Manager at Canopy Biosciences specializing in ChipCytometry™, the Precise Spatial Multiplexing platform enabling single-cell quantitative measures of virtually unlimited protein biomarkers in tissue samples and cell suspensions.

Low-cost WGS-based detection and monitoring of circulating tumor DNA using mostly natural sequencing by synthesis

Cancer Omics

POSTER 111

Itai Rusinek[1], **Eti Meiri**[1], **Omer Barad**[1], **Doron Lipson**[1], **Christopher G. Smith**[2,3], **Pippa G. Corrie**[3,4], **Nitzan Rosenfeld**[2,3]

1 Ultima Genomics, Newark, CA
2 Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, United Kingdom
3 Cancer Research UK Cambridge Centre, Cambridge, United Kingdom
4 Department of Oncology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom

Detection of circulating tumor DNA (ctDNA) in cell-free DNA can help identify patients who will likely recur and monitor treatment efficacy based on detection and quantification of residual ctDNA during and after treatment. Currently available clinical assays for residual disease employ deep sequencing of a limited number of targeted mutations or methylation markers. An alternative approach, for overcoming sampling limitations for limited samples or low ctDNA concentrations and simplifying the clinical workflow, employs whole genome sequencing (WGS) to detect somatic mutations in cell-free DNA across the genome. Proof of concept studies and statistical models suggest a limit of detection (LoD) of <10-5 can be achieved using this approach, limited only by the affordable depth of WGS and the sequencing noise.

Here we present the application of a novel sequencing platform employing mostly-natural sequencing by synthesis (mnSBS) chemistry for detection and quantification of ctDNA using WGS. mnSBS provides two distinct advantages in this setting. First, the ability to generate low-cost WGS data at scale ($1/Gb) enables affordable implementation even at greater WGS depth. Second, since mnSBS involves measuring a single nucleotide type in each cycle, it can be exquisitely specific for single nucleotide substitutions which comprise the majority of somatic mutations in most cancers.

We demonstrate this application by deep WGS of 25 samples from 5 melanoma patients including matched tumor/normal and 1-3 plasma cell-free DNA samples from each patient. A list of somatic mutations was generated for each patient by comparing the tissue sample (average sequencing depth of 111x) with the matched normal samples (39x) and filtered to include only variants that are unlikely to be sequencing error, resulting in 6k-86k mutations per patient at a mean tumor allele fraction of 13%-50% for the different patients. Testing all 15 plasma samples (129x) for the 5 different mutation lists by WGS, we identified the respective mutations in all patient-matched samples at tumor fractions ranging from 0.08% to 33%, with 0.96 correlation to ctDNA fractions measured by a targeted ctDNA assay (InVisionSeq®, Inivata Ltd.). Background levels of the mutations in unmatched samples were measured at 4x10-5, with evidence suggesting that most of the error originates from sample preparation and can likely be reduced by further optimization.

The ability of mnSBS to accurately quantify the level of ctDNA by WGS suggests this method could enable development of affordable cancer monitoring assays with high analytical sensitivity that is required for clinical applications.

Spatially mapped quantitative single-cell characterization of PD-L1–related RNA microenvironments in triple-negative breast cancer

Cancer Omics

POSTER 112

Jodi M. Carter[1], **Sean Kim**[2], **Mark Gregory**[2], **Nathan Schurman**[2], **Youngmi Kim**[2], **Joseph Beechem**[2], **E. Aubrey Thompson**[3]

1 Laboratory Medicine and Pathology, Mayo Clinic, Rochester MN
2 NanoString® Technologies, Seattle WA
3 Department of Cancer Biology, Mayo Clinic, Jacksonville, FL

Triple-negative breast cancer (TNBC) is an aggressive disease with limited therapeutic options. PD-L1-targeted checkpoint blockade has shown efficacy in a subset of patients, but therapeutic response prediction's immune and other biomarker correlates are not well understood. NanoString's new Spatial Molecular Imaging (SMI) platform is designed to generate high-plex, single-cell quantitative RNA profiles with preservation of spatial context. In this study, we used SMI technology to analyze the single single-cell spatial transcriptomic landscape of TNBC. Specifically, our initial goal was to identify and map PD-L1+ cells (verified with clinical FDA-approved PD-L1 companion assays) within user-defined fields of view, to characterize differences between PD-L1+ and PD-L1- immune and/or epithelial cells, and to elucidate molecular features associated with interactions between such cells and neighboring cells within the tumor microenvironments. FFPE tissue biopsies taken at diagnosis or disease progression were profiled using SMI with a 1000-plex immuno-oncology RNA panel. The results of these studies provide fundamental new insight into the immune architecture of TNBC tumors, and such insight into the spatial immune landscape of TNBC may inform a rational approach to (combinatorial) immunotherapy in patients with TNBC.

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Digital microfluidics–enabled whole exome sequencing automation enhances next generation sequencing (NGS) clinical applications

Cancer Omics

POSTER 113

Phillip Michaels[1,2], **Usman A. Tahir**[2,3], **Athanasios Ploumakis**[1,2,4,5], **Eugenia Carvajal**[6], **Preetham Hosur**[6], **Eduardo Cervantes**[6], **Winston Hide**[1,2,4], **Ioannis S. Vlachos**[1,2,4,5]

1 Department of Pathology, Beth Israel Deaconess Medical Center, Boston, MA
2 Harvard Medical School, Boston, MA
3 Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA
4 Harvard Medical School Initiative for RNA Medicine, Boston, MA
5 Broad Institute of MIT and Harvard, Cambridge, MA
6 Miroculus, Inc., San Francisco, CA

Next generation sequencing (NGS)-based molecular diagnostics have revolutionized patient management and stratification. Technological advancements enabled a constant reduction in cost and turnaround time, while increasing accuracy, targeted regions, and throughput in sequencing somatic and germline events to efficiently inform diagnosis, prognosis, and treatment strategies. BIDMC's Genomic Medicine Initiative is evaluating the routine clinical use of cutting-edge approaches, including whole exome sequencing (WES) as part of a routine clinical and diagnostic practice. Clinical settings require robust and reproducible assay performance, with lab automation solutions being an optimal avenue to address this goal. However, liquid handling automation performs best when high numbers of samples are batched together. As an alternative "on-demand" automation solution, we evaluated the use of a programmable digital microfluidic system (Miroculus Miro Canvas), for the development of micro-scale lab processes including amplified NGS libraries.

As part of our evaluation, we generated replicate libraries (eight total) of the Genome in a Bottle (GIAB) consortium NA12878 reference DNA using 50 ng as input into Twist enzymatic fragmentation library preparation workflow. Library yields, after on-board PCR (10 cycles), ranged from 100 to 160ng/uL with size distributions between 375 and 420 bp. Library enrichment was performed using the Twist fast hybridization target enrichment kit with the Human Core Exome panel. Pooled post-enrichment PCR required 8 cycles to produce a final 125 nM library pool that was sequenced using a NextSeq High output PE150 run. 90% of targeted regions were covered at 30X when downsampling to 35M reads. Compared to manual methods, libraries made using the digital microfluidics instrument are equivalent or better in terms of yield, data quality, reproducibility, and turnaround time.

We also evaluated the potential to apply the Miro Canvas in a clinical molecular diagnostics setting to combine NGS library prep and enrichment for fresh frozen and FFPE samples. The system demonstrated the potential to reduce library preparation scale, increase flexibility, and to automate clinical workflows, while enabling high repeatability and robustness.

An unprecedented longitudinal view of the mutational and transcriptomic landscape of an aggressive chemonaive sarcoma in a patient with exceptional survival

Cancer Omics

POSTER 114

Deep Pandya[1], **Jean-Noël Billaud**[2], **Boris Reva**[3], **Kelsey Engelman**[3], **Maria Mercedes Padron**[3], **Raja M. Flores**[4], **Jamal Rahaman**[5], **John A. Martignetti**[1,3], **Tatiana Omelchenko**[6]

1 The Rudy L. Ruggles Biomedical Research Institute, Nuvance Health, Danbury, CT
2 QIAGEN Bioinformatics, Redwood City, CA
3 Genetics and Genomic Science, Icahn School of Medicine at Mount Sinai, New York, NY
4 Department of Thoracic Surgery, Icahn School of Medicine at Mount Sinai, New York, NY
5 Division of Gynecologic Oncology, Department of Obstetrics, Gynecology and Reproductive Science, Mount Sinai Medical Center, New York, NY
6 Laboratory of Mammalian Cell Biology and Development, the Rockefeller University, New York, NY

Introduction: Uterine leiomyosarcomas (uLMS) are a rare aggressive sarcoma of the smooth muscle layer of the uterus, clinically demonstrating both high local and metastatic recurrence rates. Effective treatment remains challenging. Poor understanding of the molecular mechanisms driving uLMS growth and spread makes it difficult to target the disease using current chemo-, radio- and immunotherapies. To gain deeper insight into genetic drivers and possible targeted treatments, we deeply genomically characterized an aggressive metastatic uLMS from a patient with long-term survival managed solely by surgery alone. This treatment naïve uLMS offers a rare insight into metastatic evolution without bias from chemotherapy.

Methods: We describe the case of a 55 yo female, a treatment-naive patient with aggressive metastatic uLMS with multiple pulmonary metastases who has experienced long-term survival for >7 years without therapy except surgery. In total, 5 tumor (1 primary and 4 lung metastasis) and 2 normal lung adjacent tissue were analyzed by whole exome sequencing (WES) and RNAseq (NEBNext Ultra II). Bioinformatics analysis was performed using QIAGEN CLC Genomics Workbench, Ingenuity Pathway Analysis and Clinical Insight tools.

Results: Using this unique longitudinal sampling, we identified shared and unique mutational signatures (Figure 1A) and transcriptional expression signatures across time of a primary tumor and its metastases (Figure 1B). WES identified 79 shared mutations and 299 copy number variations (CNV). Analysis of CNV revealed biallelic loss of TP53, RB1 and a somatic PTEN variant with single copy loss of the wild type allele. Transcriptomic analysis depicted pathways involved in hormone and receptor signaling regulating cell proliferation, the cytoskeleton and cell migration, myogenic de-differentiation, and inflammation. Consistent with genomic alterations, the PTEN/PI3KCA/AKT/mTOR pathway was identified as an important and of clinically actionable candidate therapeutic target.

Conclusions: We identified mutations and pathways as potential drivers of tumor growth and we propose mechanisms and targeted therapeutic strategies for this uterine leiomyosarcoma patient that are applicable to other sarcomas with specific genomic alterations.

Cancer Omics

POSTER 115

The spatial landscape of copy number variation in benign and malignant tissue

Andrew Erickson[1], **Emelie Berglund**[2], **Mengxiao He**[2], **Maja Marklund**[2], **Reza Mirzazadeh**[2], **Niklas Schultz**[3], **Linda Kvastad**[2], **Alma Andersson**[2], **Ludvig Bergenstråhle**[2], **Joseph Bergenstråhle**[2], **Ludvig Larsson**[2], **Alia Shamikh**[4,5], **Elisa Basmaci**[4,5], **Teresita Diaz De Ståhl**[4,5], **Timothy Rajakumar**[1], **Kim Thrane**[2], **Andrew L Ji**[6], **Paul A Khavari**[6], **Firaz Tarish**[3], **Anna Tanoglidi**[7], **Jonas Maaskola**[2], **Richard Colling**[1,8], **Tuomas Mirtti**[9,10,11], **Freddie C Hamdy**[1,12], **Dan J Woodcock**[1,13], **Thomas Helleday**[3,14], **Ian G Mills**[1], **Alastair D Lamb**[1,12], **Joakim Lundeberg**[2]

1 Nuffield Department of Surgical Sciences, University of Oxford, United Kingdom
2 Department of Gene Technology, KTH Royal Institute of Technology, Science for Life Laboratory, Solna, Sweden
3 Science for Life Laboratory, Department of Oncology-Pathology, Karolinska Institutet, Solna, Sweden
4 Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden
5 Department of Clinical Pathology and Cytology, Karolinska University Hospital, Stockholm, Sweden
6 Program in Epithelial Biology, Stanford University School of Medicine, Stanford, CA
7 Department of Clinical Pathology, University Uppsala Hospital, Uppsala, Sweden
8 Department of Cellular Pathology, Oxford University Hospitals NHS Foundation Trust, United Kingdom
9 Department of Pathology, University of Helsinki & Helsinki University Hospital, Helsinki, Finland
10 Research Program in Systems Oncology, Faculty of Medicine, University of Helsinki, Helsinki, Finland
11 iCAN-Digital Precision Cancer Medicine Flagship, Helsinki, Finland
12 Department of Urology, Oxford University Hospitals NHS Foundation Trust, United Kingdom
13 Big Data Institute, Old Road Campus, University of Oxford, United Kingdom
14 Weston Park Cancer Centre, Department of Oncology and Metabolism, University of Sheffield, United Kingdom

Defining the transition from benign to malignant tissue is fundamental to improve early diagnosis of cancer. Gene expression has previously been used to infer CNVs in single cells, successfully identifying regions of chromosomal gain and loss. We expand this into a spatial modality and infer genome-wide CNV from spatially resolved mRNA profiles in situ to study spatial genome integrity. This was applied to a cross section of an entire prostate organ to explore the spatial iCNV landscape of a commonly multifocal malignancy. Using spatial transcriptomics we obtained organ-wide transcriptional information from over 21 000 barcoded regions with a mean average of 3500 expressed genes detected per barcoded spot. Using factorized negative binomial regression, we found 25 factors showing overlap between histology and gene expression factors representing tumor, hyperplasia and benign epithelial. Interestingly, spatial iCNV identified regions with increased iCNV activity distinct from morphology or expression analysis.

(continue to page 39)

Cancer Omics

POSTER 115

(continued from page 38)

To increase the fidelity of our analysis of variable iCNV regions we took advantage of smaller barcoded spots (Visium, 10x Genomics) to perform a more detailed interrogation of key sections. Analysis of data from approximately 30 000 sots showed that while certain gene expression factors displayed a fairly homogenous inferred genotype, others were strikingly heterogenous. We constructed a clone-tree to describe sequential clonal events versus independently arising cancer-clone and interestingly areas of spatially homogeneous Gleason pattern showed evidence of clonally distributed CN heterogeneity. Many CNVs seem to also occur in benign tissue, most notably in chromosome 8, which has been well-described in aggressive prostate cancer, but also several other CN gains and losses. The findings were validated through an additional 37 000 spots from a cross-section of another prostatectomy which confirmed the spatial continuum of benign clones in proximity to cancer with shared truncal events. We also confirmed the heterogeneity of iCNV clones within prostate tumour loci, and the presence of key somatic events in benign prostate glands. We then generalized our findings in multiple organs.

In total we employed spatially resolved transcriptomics to infer spatial copy number variations in >120 000 regions across multiple organs, in benign and malignant tissues. We demonstrate that genome-wide copy number variation reveals distinct clonal patterns within tumors and in nearby benign tissue using an organ-wide approach focused on the prostate. Our results suggest a model for how genomic instability arises in histologically benign tissue that may represent early events in cancer evolution.

Comprehensive analysis of exitron splicing in human cancer

Cancer Omics

POSTER 117

Rendong Yang[1,2], **Tingyou Wang**[1]

1 The Hormel Institute, University of Minnesota, Austin, MN
2 Masonic Cancer Center, University of Minnesota, Minneapolis, MN

Background: Alternative splicing (AS) of pre-mRNA plays an important role in generating functional diversity of genes and is frequently observed in cancers. The advances of next-generation sequencing technology allowed researchers to characterize AS in a genome-wide scale and identify unconventional splicing mechanisms that deviate from the well-defined rules of splicing. Recently, a new type of AS events termed exitron (exonic intron) was discovered in plant and human genomes. Exitrons comprise protein-coding intron sequences within protein-coding exons and aberrant exitron splicing was found to play an important role in tumorigenesis and disease progression in breast cancer. However, the functional role of exitron for other cancer types is still unknown.

Methods: We developed novel computational approaches for identification and quantification of exitron splicing events and performed integrative multi-omics pan-cancer analysis of large tissue and cell line samples.

Results: we explored exitron splicing events across 33 The Cancer Genome Atlas (TCGA) cancer types from 9,600 patients by reanalyzing genomic, transcriptomic and proteomic data. We found exitron splicing events were significantly enriched in tumors compared with normal tissues. Remarkably, we have found tumor specific exitron splicing are enriched in known cancer driver genes. Additionally, we found high tumor-specific exitron burden was predictor of poor overall survival in multiple cancers. Gene set enrichment analysis on the top vs. bottom quartile patients by exitron load across TCGA cohorts revealed a significant enrichment in cell cycle pathways (mitotic spindle, G2M checkpoint), and metabolic pathways (oxidative phosphorylation). Comparison between tumor-specific exitrons and somatic mutations indicated that their occurrence are mutually exclusive suggesting exitrons could represent independent oncogenic processes. Finally, exitron-derived neoantigen burden was discovered associated with response to immune checkpoint inhibitor therapy in clear cell renal cell carcinoma, indicating that exitron-derived neoantigens could be highly immunogenic.

Conclusions: Our study demonstrates that a) Exitron splicing is a relatively unknown class of RNA level alterations that are prevalent and dysregulated in tumors compared with normal tissues; b) Exitrons splicing is a complementary source of protein-coding alterations that impact driver genes in cancers and may drive tumorigenic progress via distinct biological mechanism and pathways and c) Exitron splicing is a source of neoantigen and contributes to treatment response to cancer immunotherapy.

Conflict of Interest: None

Single-cell clarity and heterogeneity in copy number profiles in primary synovial and Ewing sarcoma with ResolveDNA™ genomic amplification

Cancer Omics

POSTER 118

Katherine A. Kennedy[1], **Jon S. Zawistowski**[1], **Viren R. Amin**[1], **Durga M. Arvapalli**[1], **Isai Salas-González**[1], **Natalie Jäger**[2], **Hanno Glimm**[3], **Gary L. Harton**[1], **Stefan Fröhling**[2], **Jason A.A. West**[1], **Stefan Pfister**[2]

1 BioSkryb Genomics, Inc., Durham, NC
2 Hopp Children's Cancer Center Heidelberg (KiTZ) and German Cancer Consortium (DKTK), Heidelberg, Germany
3 NCT Dresden and German Cancer Consortium (DKFZ), Dresden, Germany

Soft tissue sarcomas are a diverse set of mesenchymal malignancies frequently driven by translocation events, resulting in expression of aberrant fusion gene products that fuel cell proliferation and tumor growth. Examples of these, including Ewing sarcoma and synovial sarcoma, have limited data surrounding the broader genomic alterations, and intra-tumoral heterogeneity, that synergizes with and contributes to disease aggression and prognosis.

In a collaboration with the clinical, translational and basic research sarcoma expertise at the Hopp Children's Cancer Center Heidelberg (KiTZ), we utilized ResolveDNA™ kit to amplify the genomes of individual sarcoma cells with an aim to resolve copy number aberration (CNA) at the single cell level in primary sarcoma samples. The ResolveDNA™ chemistry attenuates amplicon size, redirecting the amplification primer to the primary DNA template of interest and avoiding exponential copying of amplicons. This results in unprecedented genomic coverage and uniformity, with high SNV precision and sensitivity, allelic balance, and the enhanced ability to accurately call CNA.

Our initial study comprised cells from two patient samples: a putative synovial sarcoma and a Ewing sarcoma that were FACS sorted as single cells in each well. We defined CNA using the Ginkgo algorithm and identified discrete profiles both between and within sarcoma samples. The vast majority of putative synovial sarcoma single cells displayed a highly disrupted genome, with stark focal, sub-chromosomal gains and losses suggestive of chromothripsis. While rare, isolated cells demonstrated a lack of any apparent CNA. Intriguingly, single-cell heterogeneity was notable in multiple sub-chromosomal regions including 4p and 16q. An SS18:SSX translocation was not identified in this patient, suggesting that these cells may have originated from a secondary sarcoma as opposed to a synovial sarcoma. In contrast, the Ewing sarcoma single cells harbored discrete, full-chromosomal gains and losses. Concurrently, we have extended these findings to catalog alterations at the single nucleotide level using whole exome sequencing (WES) from the same single cells that were profiled by CNV. We also analyzed primary samples from additional sarcoma subtypes to define genomic evolution and clonal lineage contributing to these pathologies.

The heterogeneity observed in this study demonstrates the power of single cell vs bulk analysis to reveal the nature of genomic instability and aid in the characterization of sarcoma subtypes at unprecedented levels of resolution. These abilities are particularly crucial to understand disease development and progression in orphan sarcomas, especially those with complex genome rearrangements and/or poor clinical outcome.

Cancer Omics

POSTER 119

EM-seq permits accurate and precise methylome analysis of cfDNA and FFPE DNA samples

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

New England Biolabs, Ipswich, MA
*Presenting author

The study of DNA methylation from cell-free DNA (cfDNA) and DNA extracted from formalin fixed paraffin embedded (FFPE) tissues are notoriously difficult. The ability to use these DNA inputs is becoming increasingly important in the field of cancer diagnostics. The cytosine modifications, 5-methylcytosines (5mC) and 5-hydroxymethylcytosines (5hmC), are key epigenetic factors that play an important role in cellular processes and their mis-regulation results in diseased states like cancer. Advances in the study of these epigenetic factors in combination with improvements in sample preparation have enabled cancer biomarker identification based on methylation profiling.

The historical gold standard for detecting cytosine methylation has been bisulfite sequencing. This method uses chemical conversion of cytosines into uracils and has been used for both targeted and whole genome methylation analysis. However, bisulfite conversion leads to DNA damage which results in shorter DNA insert sizes as well as biases in the data. It is challenging to make bisulfite libraries from cfDNA and FFPE DNA as typically the DNA is of low quantity and quality.

We developed an enzyme-based methylation detection technology (EM-seq) to address the flaws inherent to bisulfite sequencing. This method, primarily due to the intact nature of DNA after conversion, results in better quality libraries with longer insert sizes, lower duplication rates and minimal GC bias. EM-seq cfDNA and lung FFPE DNA libraries showed that the EM-seq libraries had longer inserts, lower duplication rates, higher percentages of mapped reads and less GC bias compared to bisulfite libraries. EM-seq libraries also identified a higher number of CpGs with even coverage facilitating accurate methylation calling. EM-Seq's superior sequencing metrics establish it as a standard method for robust methylation profiling, particularly for these types of challenging DNA samples.

Cancer Omics

POSTER 120

Sequence variation in mtDNA and extreme survival in glioblastoma

Sion Williams[1], **Claudine Samanic**[3], **Jamie K. Teer**[4], **Zachary J. Thompson**[4], **Jordan H. Creed**[5], **Sepideh Mokhtari**[6], **Brooke L. Fridley**[4], **L. Burt Nabors**[7], **Kathleen M. Egan**[3]

1 Sylvester Comprehensive Cancer Center; University of Miami Miller School of Medicine, Miami, FL
2 Neurology Basic Science Division, University of Miami Miller School of Medicine, Miami, FL
3 Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center, Tampa, FL
4 Department of Biostatistics and Bioinformatics, H. Lee Moffitt Cancer Center, Tampa, FL
5 Department of Health Informatics, H. Lee Moffitt Cancer Center, Tampa, FL
6 Department of Neuro-Oncology, H. Lee Moffitt Cancer Center, Tampa, FL
7 Division of Neuro Oncology, Department of Neurology, University of Alabama at Birmingham, Birmingham, AL

Malignant gliomas are the most common primary adult brain tumors, with a poor prognosis and ill-defined etiology. Strategies to exploit the high energy demands in the brain and dependency of tumors on glycolysis have been proposed as an adjunct to current therapy for glioblastoma multiforme (GBM). Mitochondria fuel most of the energy demands of the body through oxidative phosphorylation (OXPHOS) and play a central role in energy reprogramming in the cell, a long-standing 'hallmark of cancer'. A target of emerging interest in glioma research is the mitochondrial genome (mtDNA), an independent 16.6 kb multi-copy component of the genome that encodes subunits of the OXPHOS. Mitochondrial DNA (mtDNA) sequence variants and haplogroups have been linked with certain cancers, but research on GBM is lacking. We have sequenced the germline mtDNA from 420 glioma cases from the GliomaSE cohort, a large, multicenter epidemiologic study of glioma conducted at referral centers in the southeastern US. Cases were selected to include a subset with survivorship of three years or greater from diagnosis. Initial analysis of inherited variants by haplotype and gene function has not demonstrated associations with GBM survival. As mtDNA is also vulnerable to acquiring somatic mutations we are investigating somatic mtDNA mutations in 39 tumors with matched germline mtDNA sequence data using TwinStrand duplex sequencing. This approach enables variant allele detection within the physiologic range of 10-4 to 10-5 per base. Based on previous work examining mtDNA somatic variant load in normal aging we anticipate this improved approach will reveal the somatic variant spectrum of mtDNA in the gliomas under study and anticipate it will inform larger scale validation studies.

Cancer Omics

POSTER 121

DNA methylation signatures in solid and liquid biopsies

Gabriel E. Rech[1], **Rachel Goldfeder**[1], **Rahul Maurya**[1], **Alexey V. Danilov**[2], **Chia-Lin Wei**[1]

1 The Jackson Laboratory for Genomic Medicine, Farmington, CT
2 City of Hope National Medical Center, Duarte, CA

Liquid biopsies are one of the most exciting and promising applications in genomic medicine. Our ability to detect tumor-related genomic alterations amongst plasma cell free DNA (cfDNA) has opened a world of possibilities in the search for a minimally invasive, cost-efficient, and clinically convenient cancer diagnostic and prognostic assay. Nevertheless, looking for tumor-related somatic mutations in cfDNA poses several challenges like low mutation frequency and limited information about the tissue origins of the cancerous sequences. Epigenetic modifications, specially 5-Hydroxymethylcytosine (5hmC) of DNA, have been recently proposed as a more valuable biomarker for a variety of human cancers and other complex diseases. This is mainly because, even when the precise functions of 5hmC modifications remain unknown, 5hmc patterns have already been shown to have a tissue-specific distribution and unique patterns in tumor DNA when comparing with normal tissue.

In the present pilot study, we attempted to understand on the genomic features of 5hmc associate with cancerous tissues and evaluate their potential as biomarkers in patient plasma. We used high coverage bisulfite and oxidative bisulfite whole genome sequencing to characterize 5hmc methylation patterns at the single base resolution in two tissue types and their corresponding cancer pairs: normal urothelium tissue / bladder cancer and normal b-cells / lymphoma. When comparing normal tissues, we identified hundreds of differentially 5hmc methylated genes (DMGs) and enhancers (DMEs) that were in fact significantly enriched with the corresponding tissue specific functions and structures. Moreover, we found genes and enhancers with elevated 5hmc methylation levels positively correlate with gene expression levels. We then explored 5hmc patterns in normal tissue when comparing with its corresponding cancer samples and found several DMGs and DMEs as being cancer-specific biomarkers, including some already well-known genes implicated in cancer such as EGFR, MYH9, TFRC and CREBBP.

Finally, we investigated whether cfDNA is a faithful surrogate for tumor biopsy when accessing tumor specific methylation patterns. To do that, we analyzed plasma cfDNA samples from patients with bladder cancer and lymphoma and explored whether 5hmc methylation recapitulated the patterns we observed for tissues. Overall, in this pilot study we were able to 1) better understand genome-wide 5hmc methylation distributions in two tissue types and their associated cancers; 2) identify several genomic regions candidates to be assessed as potential cancer-specific biomarkers and 3) evaluate these targets directly on cfDNA samples from cancer patient.

Cancer Omics

POSTER 122

Efficient single-cell sequencing for full-length transcripts to decipher subclonal structure of melanoma

Charli Gruen[1*], **Antonis Kokkalis**[2], **Kerrie Marie**[1], **Christina Marcelus**[1], **Farid Rashidi Mehrabadi**[3], **Salem Malikić**[3], **Michael C. Kelly**[4], **Cari Smith**[5], **Sung Chin**[5], **S. Cenk Sahinalp**[3], **Chi-Ping Day**[1]

1 Laboratory of Cancer Biology and Genetics, National Cancer Institute, Bethesda, MD
2 seqWell, Inc., Beverly, MA
3 Cancer Data Science Laboratory, National Cancer Institute, Bethesda, MD
4 Single Cell Analysis Facility, Frederick National Laboratory, Bethesda, MD
5 Laboratory Animal Science Program, Frederick National Laboratory, Bethesda, MD
*Presenting Author

Intratumoral heterogeneity is driven by expansion of mutant subclones, each with a distinct metastatic potential and ability to evade therapy. Therefore, subclonal evolution has a profound impact on therapeutic outcomes. In principle, subclones can be characterized and traced by phylogenetic analysis of mutations in single cells. Unfortunately, single-cell full-length transcript sequencing (scRNA seq) datasets that allow for sufficient mutation calling are sparce. In this study, we isolated single cells from the genetically heterogeneous and phenotypically diverse “M4” mouse melanoma (Perez-Guijarro et al., Nature Med 2020) in vivo and subjected them to preparation of full-length transcript cDNA libraries following Smart-seq2 protocol and plexWell Rapid single cell protocol from seqWell workflows. When sequenced to reach one million reads per cell, both protocols generated comparable results, including cell percentage of mitochondrial genes and number of genes detected. However, the plexWell Rapid protocol yielded a significantly higher turnover rate (~50%) and lower labor demand, making it a sensitive and powerful profiling of transcript at single-cell resolution method. We employed a novel tumor progression tree inference method, Trisicell (Rashidi Mehrabadi et al., bioRxiv, 2021), to reconstruct the entire lineage history of the melanoma cells based on expressed mutations derived from scRNAseq data. We demonstrated that Trisicell accurately generates phylogenies based on scRNA mutations, allowing the tracing of subclonal evolution in tumors, thus providing a tool that can be leveraged in the clinic to uncover mechanisms of tumor cell evolution. Our results showed that the trees generated from the Smart-seq2 and plexWell Rapid datasets have corresponding subclonal structures. Interestingly, each subclone has a distinct developmental state and sensitivity to immunotherapy. Moreover, our results suggest that each subclone could develop a unique resistance mechanism, causing disease recurrence. This study highlights the importance of using scRNAseq to investigate tumor cell evolution and the benefits of using efficient library preparation method to generate high-quality datasets.

Regional expression profiling of rare tumor cells and tumor microenvironment populations in classic Hodgkin's lymphoma

Cancer Omics

POSTER 123

Tomohiro Aoki[1,2], **Kit Fuhrman**[3], **Aixiang Jiang**[1,2], **Susana Ben-Neriah**[1], **Pedro Farinha**[1,2], **Christian Steidl**[1]

1 Centre for Lymphoid Cancer, BC Cancer, Vancouver, British Columbia, Canada
2 Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia, Canada
3 NanoString Technologies, Inc., Seattle, WA

Classic Hodgkin lymphoma (CHL) is a cancer derived from B lymphocytes and typically occurs in young adults. CHL is unique amongst virtually all cancers as the malignant Hodgkin and Reed Sternberg (HRS) cells are vastly outnumbered by reactive, non-neoplastic cells in the tumor microenvironment (TME). HRS cells are typically found at a low cellular frequency in tissue biopsies ranging from 0.1 to 1% and are typically in close contact with and surrounded by T cells and macrophages. Although prior studies have demonstrated the importance of cellular crosstalk in the TME, precise phenotyping of HRS cells and surrounding immune cells remained technically very challenging. In particular, gene expression studies leveraging the resource of routinely available formalin-fixed, paraffin-embedded tissues are sporadic. Here, we sought to overcome significant limitations of gene expression analysis in HL using protein marker-guided regional profiling of HRS cells and TME components using a GeoMx® Digital Spatial Profiler.

We performed GeoMxTM Human Whole Transcriptome Atlas analysis on a tissue microarray of CHL samples (n= 19). In total, 48 areas of interest (AOI) were profiled on a GeoMxTM Digital Spatial Profiler. CD30, CD3, and CD68 were used as morphological markers to identify HRS cells, T cells, and macrophages, respectively. Differential expression analysis was performed between areas of high CD30, CD3, and CD68 expression to confirm effective transcriptome enrichment in each area.

Differential expression analyses between CD30+ regions and CD3+ regions identified CCL17 (encoding TARC), a known HRS cell marker, as significantly upregulated in CD30+ AOIs compared to CD3+ AOIs. Furthermore, gene set enrichment analysis revealed that CD30+ AOIs were significantly enriched for an HRS signature gene set defined by Affymetrix gene expression data generated from laser microdissected HRS cells.We confirm the feasibility and applicability of regional gene expression profiling using a GeoMxTM Digital Spatial Profiler to phenotype HRS cells in the context of histology sections of formalin-fixed paraffin-embedded tissues.

Launch of a ~$100 clinical whole genome assay for quantification of tumor fraction in circulating cell-free DNA

Cancer Omics

POSTER 124

Micah Rickles-Young[1], **Carrie Cibulskis**[1], **Junko Tsuji**[1], **Alyssa Macbeth**[1], **Gavin Ha**[4], **Caroline Petersen**[1], **Maegen Harden**[1], **Katie Larkin**[1], **Gina Vicente**[1], **Nicole Francis**[1], **Stacey Gabriel**[1], **Viktor Adalsteinsson**[1,3], **Dan Stover**[1,2], **Niall Lennon**[1]

1 Broad Institute of MIT and Harvard, Cambridge, MA
2 Ohio State University, Columbus, OH
3 Koch Institute of MIT, Cambridge, MA
4 Fred Hutchinson Cancer Research Center, Seattle, WA

Minimally-invasive genomic analysis of cell-free DNA (cfDNA) to guide rational therapeutic decisions has the potential to transform the treatment of cancer. Work by our group and others in metastatic breast cancer demonstrated that tumor fraction is strongly associated with metastatic survival. Additionally, rapid changes in tumor fraction (TFx) early in the course of breast cancer treatment are associated with progression-free survival. Recognizing the importance of %cfDNA TFx estimation as a promising prognostic and predictive biomarker in metastatic breast and other cancers, our lab has clinically validated and launched an efficient shallow-depth, whole genome assay to estimate TFx.

Our ultra-low pass whole genome (ULP-WGS, 0.1x depth) cfDNA assay and associated computational pipeline (ichorCNA) allows rapid and precise quantitation of tumor DNA content in plasma as well as discovery of copy number alterations from a single blood sample. Building on our experience in processing > 10,000 cfDNA specimens in our research laboratory, we have now completed a full CLIA/CAP performance evaluation and are launching the assay for clinical research initiatives. The ULP-WGS TFx assay achieves 100% sensitivity to detect ctDNA >= 3% expected tumor fraction, with 97.14% specificity. Furthermore, the ichorCNA tumor fraction results were highly consistent in both repeatability and reproducibility assessments of 60 patient specimens. We will present the full technical methodology, validation strategy, and results of our in-depth clinical performance evaluation. Furthermore, we will present initial plans for deploying this assay in future clinical trials.

Informatics and Computational Biology

Informatics and Computational Biology

POSTER 200

A community-driven roadmap to advance research on translated open reading frames discovered by Ribo-seq

Jonathan M. Mudge[1], **Jorge Ruiz-Orera**[2], **John Prensner**[3], **Sebastiaan van Heesch**[4]

1 European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom
2 Cardiovascular and Metabolic Sciences, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany
3 Broad Institute of Harvard and MIT, Cambridge, MA
4 Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands

Ribosome profiling (Ribo-seq) has catalyzed a paradigm shift in our understanding of the translational 'vocabulary' of the human genome, discovering thousands of translated open reading frames (ORFs) within long non-coding RNAs and presumed untranslated regions of protein-coding genes. Nonetheless, 'Ribo-seq ORFs' are 'non-canonical' at present; they are not incorporated into reference gene annotation catalogs, and remain effectively 'invisible' to both large-scale scientific projects and variant interpretation workflows in the clinic.

Alongside technical considerations, the lag in annotating Ribo-seq ORFs is largely down to uncertainties on their biological roles, with the general lack of evolutionary constraint on their coding sequences thwarting obvious functional interpretation. Nonetheless, it is indisputable that at least some Ribo-seq ORFs make stable proteins. Others are known to mediate gene regulation, and Ribo-seq has enormous potential to identify 'uORFs' that control protein translation in medically important genes. Furthermore, Ribo-seq may help shed light on a longstanding blindspot in gene annotation: the preponderance of translation events that are specific to disease states. This may be especially true for cancer, where a developed understanding of non-canonical translation may become important in diagnostics and perhaps even therapeutics.

Here, we outline an international community-led effort to produce an initial consolidated and standardized catalog of 7,264 human Ribo-seq ORFs, unifying the efforts of fourteen research groups and database projects. This includes Ensembl / GENCODE alongside UniProt and HGNC, which makes this the point at which Ribo-seq ORFs enter the mainstream annotation 'ecosystem'. Nonetheless, these are the first steps of a long-term endeavour, and – as we shall discuss – there are many knowledge gaps that require urgent attention. Answers will support the development of more accurate annotations and variant interpretations, with the potential to yield substantial insights across all aspects of human biology.

Informatics and Computational Biology

POSTER 201

Optimizing the GATK for the Ultima Genomics sequencing platform: an evaluation of 200 whole genome HapMap samples

Megan Shand[1], **Ilya Soifer**[2], **Yossi Farjoun**[1,3], **Omer Barad**[2], **James Emery**[1], **Takuto Sato**[1], **Ricky Magner**[1], **Matthew Coole**[1], **Tom Howd**[1], **Stacey Gabriel**[1], **Eric Banks**[1]

1 Broad Institute of MIT and Harvard, Cambridge, MA
2 Ultima Genomics, Newark, CA
3 Lady Davis Institute for Medical Research, Montreal, Canada

Here we present optimizations and updates to the Broad Institute's Genome Analysis Toolkit (GATK) that were used to call germline short variants from single-read, flow-based data from the Ultima Genomics sequencing platform. We outline efforts to reduce the cost and increase accuracy of the analysis pipeline and present an overall analysis of 200 well-characterized HapMap and Genome-In-a-Bottle (GIAB) whole genome samples.

In order to achieve high sensitivity and precision for short variant calling, we developed an updated version of the GATK and analysis workflow to account for the single-read, flow-based nature of the Ultima Genomics sequencing systems. For example, MarkDuplicates has been adapted to handle ambiguity in the aligned start of reads, and contamination estimation has been adapted to use only the highest quality SNP sites based on flow cycles and local realignment. We added features to HaplotypeCaller, the GATK variant calling tool, to handle flow-based data: the classic Hidden Markov Model (HMM) was replaced by a flow-based likelihood model that more accurately accounts for sequencing errors present in the data; additionally, candidate haplotypes are filtered by quality and strand bias. These changes significantly improve accuracy as well as decrease the computational cost of running the pipeline. All new and updated software tools will be released in an upcoming version of the open source GATK on github, and featured workspaces will be available in the Terra analysis platform.

Performance of GATK variant calling on Ultima Genomics data is similar in quality to Illumina NovaSeq6000 on SNPs over the exome regions. We observe high quality SNP and non-homopolymer INDEL calls over the High Complexity Region, with lower quality over the more difficult regions of the genome as expected. Homopolymer INDELs of less than length 8 are similarly of high quality, whereas homopolymer INDELs larger than this length deteriorate in quality and currently are not called at lengths above 12. We show results for all GIAB samples and additionally we have generated a full 200 sample callset of HapMap samples, all of which will be publicly available for download.

Informatics and Computational Biology

POSTER 202

DeepConsensus: gap-aware sequence transformers for sequence correction

Gunjan Baid[1*], **Daniel E. Cook**[1*], **Kishwar Shafin**[1], **Taedong Yun**[1], **Felipe Llinares-López**[1], **Quentin Berthet**[1], **Aaron M. Wenger**[2], **William J. Rowell**[2], **Maria Nattestad**[1], **Howard Yang**[1], **Alexey Kolesnikov**[1], **Armin Töpfer**[2], **Waleed Ammar**[1], **Jean-Philippe Vert**[1], **Ashish Vaswani**[1], **Cory Y. McLean**[1], **Pi-Chuan Chang**[1^], **Andrew Carroll**[1^]

1 Google LLC, Mountain View, CA
2 Pacific Biosciences (PacBio), Menlo Park, CA
*These authors contributed equally
^These authors contributed equally.

Pacific BioScience (PacBio) circular consensus sequencing (CCS) generates long (10-25 kb), accurate "HiFi" reads by combining serial observations of a DNA molecule into a consensus sequence. The standard approach to consensus generation uses a hidden Markov model (pbccs). Here, we introduce DeepConsensus, a deep-learning based approach for correcting errors in HiFi sequence data. DeepConsensus uses a novel alignment-based loss to train a gap-aware transformer-encoder (GATE) for sequence correction.

DeepConsensus reduces errors in PacBio HiFi reads by 41.9% compared to pbccs in human sequence data. We stratify performance across mismatches, homopolymer insertions and deletions, and non-homopolymer insertions and deletions, and DeepConsensus improves accuracy in each category. We demonstrate that using reads from DeepConsensus improves the contiguity, completeness, and correctness of genome assembly when compared to assemblies generated using pbccs reads. With two SMRT Cells of HG003, reads from DeepConsensus improve hifiasm assembly contiguity (NG50 4.9Mb to 17.2Mb), increase gene completeness (94% to 97%), reduce false gene duplication rate (1.1% to 0.5%), and improve assembly base accuracy (Q43 to Q45).

Similarly, we demonstrate improved accuracy of variant calling when using DeepConsensus reads. With two SMRT Cells of HG003, DeepConsensus reduces variant calling errors by 24% over pbccs by using DeepVariant retrained for DeepConsensus reads.

Finally, we demonstrate that improvements in accuracy allow for longer PacBio read lengths while retaining acceptable read accuracy, enabling improvements in contiguity of genome assembly and increasing the experimental design options for PacBio sequencing.

Informatics and Computational Biology

POSTER 203

Quantification and correlative analysis of imperfect dsDNA break repair

Dominika Borek[1], **Raquel Bromberg**[1], **Yirui Guo**[1], **Mahitha Roy**[1], **Zbyszek Otwinowski**[1]

1 UT Southwestern Medical Center at Dallas, Dallas, TX

In many cancers, including breast, ovarian, prostate, and pancreatic cancers, cancer cells have defective dsDNA break repair due to some dysfunction in the homologous recombination repair (HRR) pathways. HRR is the cell's highest fidelity method of repairing double-stranded DNA breaks and when it fails, e.g. due to mutations in BRCA1 and/or BRCA2 genes, DNA repair is redirected to the more error-prone mechanisms. These mechanisms may introduce errors that are not simple substitutions but have complex characteristics, distinct from replication errors, with deletions being the most common genomic scar. The deletion-containing mutational signatures have been identified before in cancer tissues, for instance in cancer cells having both copies of HRR-associated genes BRCA1 and BRCA2 inactivated. However, to be included in this mutational signature, the same deletion had to be clonally amplified, i.e. observed independently multiple times in sequencing reads. Well before a deletion is observed some arbitrary number of times in the results of sequencing, the defective HRR may generate many more deletions that happen only once or twice in all cells in an organism or in an organ. Because these somatic deletions are distributed randomly and sparsely in the genomic DNA, there are currently no efficient methods to identify them before some very small number of them becomes amplified by cancer growth.

We developed a method that identifies and quantifies these rare deletional events. The method relies on sequencing data acquired with standard approaches and doesn't require special library preparation. We tested our approach by analyzing whole genome sequencing data sets from the International Cancer Genome Consortium (ICGC). After separating signatures of non-clonal and sub-clonal deletions from signatures of experimental systematic effects, the analysis revealed patterns of correlations, which (1) stratified patients into distinct groups, (2) provided insight into somatic evolution of cancers, and (3) identified potential targets for therapeutic intervention.

Short- and long-read sequencing strategies detect numerous alternative splicing events during stem cell differentiation

Informatics and Computational Biology

POSTER 204

Dena Leshkowitz[1], **Merav Kedmi**[1], **Yael Fried**[1], **David Pilzer**[1], **Hadas Keren-Shaul**[1], **Elena Ainbinder**[1], **Bareket Dassa**[1]

1 Life Sciences Core Facilities, Weizmann Institute of Science, Rehovot, Israel

This work explored the impact of the selected sequencing platform on the detection of global changes in alternative splicing during embryonic stem cells (ESCs) differentiation. RNA was collected before and after retinoic acid-induced differentiation of ESCs into neurons, and was then sequenced with Illumina short reads (RNA-Seq), and via cDNA and Direct RNA Oxford Nanopore long-read sequencing methodologies. Long-reads technologies generate individual reads that can span the full length of transcripts. The Direct RNA approach, unlike the cDNA approach, enables the direct sequencing of individual poly-adenylated RNAs, without introducing amplification biases. This study compared both transcriptomic changes and sequencing platform differences at the gene, transcript, and exon levels. To date, this is first study comparing these sequencing platforms at the level of exon usage. Among the many genes discovered and the 22 genes validated to have differential exon usages, were key differentiation genes, such as Mdb3 and Aplp1. A collective and comprehensive resource of alternative splicing events, as revealed by the three sequencing strategies, mostly in a platform dependent manner, is provided.

Integration of single-cell multiomics and spatial gene expression in fly, mouse, and organoids

Informatics and Computational Biology

POSTER 205

Jaspar Janssens[1], **Carmen Bravo González-Blas**[1], **Nikolai Hecker**[1], **Gabriele Partel**[1], **Seppe De Winter**[1], **Joy Ismail**[1], **Stein Aerts**[1]

1 VIB-KU Leuven Center for Brain & Disease Research, Leuven, Belgium

While single-cell transcriptomics has revolutionized the genomics field by paving the way towards the discovery of novel cell types and cell-type specific regulation, a key drawback is the lack of spatial awareness. Recently, we have created an atlas of the entire fly with over 250 cell types annotated, covering cells from all major tissues. Still, a large majority of clusters remained unannotated, especially in complex organs like the brain. Knowing the location of marker gene expression of these clusters would greatly aid in the identification of cell types. Here we present spatial transcriptomic data generated using the Molecular Cartography™ from Resolve Biosciences for 100 genes in the fly head and fly body, complementing the existing fly cell atlas. We developed computational pipelines in NextFlow using both segmentation and segmentation-free approaches to integrate spatial gene expression patterns with scRNA-seq and scATAC-seq, allowing us to identify novel cell types and confirm expression of key transcription factors. For example, the transcriptomes of the central brain of the fly are largely unannotated. With the spatial maps obtained using Molecular Cartography, we were able to match different clusters with distinct gene expression patterns in the central brain, providing insight in its structure. The pipeline was then extended to an additional dataset of the mouse liver, where gradients of gene expression within hepatocytes in single-cell datasets were recapitulated as gradients in their zonation. Finally, we performed experiments on organoids and mouse brain, highlighting the usability of the system across different models. These high-quality spatial data will be complementary to existing single-cell datasets, validating marker genes and providing the missing link between cluster and cell type.

Informatics and Computational Biology

POSTER 206

Boosting GWAS sensitivity and computational efficiency with word embedding vectors

Jason Chin, Arkarachai Fungtammasan, Yih-Chii Hwang, Chiao-Feng Lin, Peter Nguyen, Daniel Quang

DNAnexus, Mountain View, CA

With the collection of a rich set of phenotypic measurements and extensive genotypic data, researchers can perform GWAS or PheWAS with the cohort from UK Biobank. The comprehensive collection of phenotype data across approximately 500,000 participants makes it possible to adapt modern machine learning techniques to improve phenotype-variant association studies. We utilized word-to-vector techniques commonly used for natural language processing with deep learning for analyzing the information derived from Electronic Health Record (EHR). A set of International Classification of Diseases version 10 (ICD10) is assigned for each participant in UK Biobank. Previous studies that used ICD10 code had identified genomic variants associated with diseases by converting the ICD10 assignments as binary traits. However, using binary traits may be too strict as some disease states may not be simple binary cases. Comorbidity conditions could provide additional information than any single ICD10 code. We identify the latent structures of the cohort with the assigned ICD10 codes by converting the ICD10 code into quantitative representations. This allows us to replace the binary traits with quantitative traits capturing the different states of the diseases with the comorbidity information. We develop a method to measure how close a participant is to a phenotype (defined by an ICD10 code) with a proxy score and perform GWAS with the proxy score.

We demonstrate that using the proxy score for GWAS may find missing variants in contrast to using simple binary traits. Applying GWAS for ICD10 code C90 (Multiple myeloma) as a binary trait with a whole-exome dataset from 200k UKB participants, we missed a variant at chr9:5073770 G>T (V617F) in one exon of JAK2. We found the variant is significantly ($p < 1e-60$) associated with the proxy score of C90. This finding re-discovered that the variant V617F was prevalent in multiple myeloma patients. To systematically evaluate this approach, we performed GWAS using Regenie for a set of total 1149 ICD codes with the derived quantitative proxy scores. We found our method was more sensitive for rare variants than using ICD10 as binary traits. Performing GWAS with the proxy score of each trait is also more efficient, about 2 to 3 times faster and cheaper than using the binary traits with Regenie. In conclusion, using word embedding techniques to convert binary traits to quantitative may improve the sensitivity of medically relevant variant discoveries and the computation efficiency for large scale GWAS and PheWAS.

Informatics and Computational Biology

POSTER 207

Automated processing of next-gen sequencing data from real-time SARS-CoV-2 surveillance

Alberto Riva[1], Massimiliano Tagliamonte[2,3], Brittany Rife-Magalis[2,3], Simone Marini[2,3], Carla Mavian[2,3], Steven J Madore[1], Marco Salemi[2,3]

1 Interdisciplinary Center for Biotechnology Research, University of Florida, Gainesville, FL
2 Emerging Pathogens Institute, University of Florida, Gainesville, FL
3 Department of Pathology, Immunology and Laboratory Medicine, University of Florida, Gainesville, FL

The ability to Determining genome sequencesprocess of SARS-CoV-2 samples quickly and accurately is critical to real-time monitoring of the progression and evolution of the current pandemic. We have developed FLACO, an integrated analysis and reporting pipeline to analyze samples from the large-scale genomic surveillance of SARS-CoV-2 cases under way at the University of Florida.

Raw genome sequencing data generated in the ICBR Next Generation Sequencing Core is first processed usingin the The first part of the analysis takes place on Illumina's Basespace cloud environment wherebyusing the . The "DRAGEN RNA Pathogen Detection" app, which maps short reads from each sample to the SARS-CoV-2 reference genome to generate , and generates a sample-specific consensus sequence. The app also produces summary reports and coverage statistics that helpto assessing the data quality. of generated data. In particular, the proportion of reads mapping to the SARS-CoV-2 genome compared to those mapping to a panel of control (human) sequences provides an indication of the viral load of the original RNA sample.'s viral load.

The dataset is then downloaded to the UF HiPerERGator a high-performance computing cluster, where further processing is performed by aa custom-developed analysis pipeline. The pipeline extracts relevant quality control information from the sample data that is stored, storing it in a local relational database. Examples of data elements collected include average depth of coverage, median depth of coverage, fraction of the viral genome covered at various depths (e.g. 20X, 50X, etc), and what fraction of the spike protein was detected. These parameters are then used to filter out samples , discarding those whose quality is insufficient for downstream analysis.

SARS-CoV-2 gGenomeic data are is also combined with metadata from the donor consisting ofincluding about the individuals that the samples were collected from. The most important metadata fields in the database are the sample collection date and location, in addition to the individual's donor age, sex, and vaccination status (if known). Consensus sequences for samples that pass our quality control criteria are then aligned to the official reference sequence, and the alignment is processed with Pangolin to generate lineage calls. All analysis results are stored in the local database, that can be queried using an ad-hoc command-line program.

(continue to page 58)

(continued from page 57)

Output of the pipeline consists of a series of tables and web-accessible reports, divided into several sections: general sequencing data, coverage information, and lineage calls. Different reports are produced for each sequencing run performed, and for each group of samples (on the basis of geographical origin). Finally, the pipeline prepares sequence and metadata files suitable for submission to GISAID in batch mode.

In six months, the pipeline has processed over 6,800 samples, 4,543 of which received a lineage. This effort has allowed us to monitor the progression of the pandemic and the dynamics of lineage spread across Florida in real time.

Primer-monitor and VarSkip primers: open access infrastructure to identify and respond to SARS-CoV-2 variant impacts on amplification methods

Matthew Campbell[1], **Gautam Naishaham**[1], **Kaylinnette Pinet**[1], **Luo Sun**[1], **Lynne Apone**[1], **Bradley W. Langhorst**[1]

1 New England Biolabs, Ipswich, MA

Molecular diagnostic technologies such as qPCR and LAMP have been essential in the detection of SARS-CoV-2 infections. Multiplex sequencing enrichment approaches have produced millions of SARS-CoV-2 sequences and enabled reliable tracking of real-time viral evolution. Both methods depend on oligonucleotide primers to amplify genomic regions of interest. Despite the relatively slow mutation rate of SARS-CoV-2 that has been observed so far, variants with impacts on transmissibility, morbidity, and detection have arisen. In some cases these variants have already been shown to impact the efficacy of specific diagnostic assays. Specific variants causing an S-gene dropout in the Thermo Fisher TaqPath assay have been reported to be associated with the B.1.1.7 variant. Negative effects on the Roche COBAS assay have also been reported. More recently, mutations associated with the Delta lineage have compromised ARTICv3 amplicons and obscured sequence results. Since variant frequencies are highly variable by geographic region, monitoring tools that reflect geographic partitioning can allow interested parties to evaluate assay effectiveness with greater precision. We have developed https://primer-monitor.neb.com/ as an open-source tool to address the ongoing need to assess potential impacts on primers as new variants emerge and used the variant frequency database to inform alternate multiplex primer designs. VarSkip primers were designed in regions with fewer variants (as of March 2021) and have proven robust in sequencing current variants of concern, with fewer variant overlaps than ARTICv3, ARTICv4 or Midnight primer sets.

A cloud-scalable software suite for proteogenomics data analysis and visualization at scale

Informatics and Computational Biology

POSTER 209

Margaret K.R. Donovan[1], **Harsharn Auluck**[1], **Arjun Vadapalli**[1], **Yan Berk**[1], **Aaron Gajadhar**[1], **Khatereh Motamedchaboki**[1], **Yuandan Lou**[1], **Asim Siddiqui**[1]

1 Seer, Inc., Redwood City, CA

Comprehensive assessment of the flow of genetic information through multi-'omic data integration can reveal the molecular consequences of genetic variation underlying human disease. Next generation sequencing (NGS) is used to identify genetic variants and characterize gene function (e.g., transcriptome and epigenome), while mass spectrometry is used to assess the proteome through characterization of protein abundances, modifications, and interactions. A new plasma profiling platform, the Proteograph Product Suite1, leverages multiple nanoparticles with distinct physiochemical properties to enable deep plasma proteome analyses at scale. Here, we present a cloud-based, data analysis software platform called Proteograph Analysis Suite (PAS) for proteogenomic data analyses through the integration of proteomics data derived from the Proteograph with genomic variant information derived from NGS experiments.

The PAS features include an experiment data management system, analysis protocols, an analysis setup wizard, and tools for reviewing and visualizing results. PAS can support both Data Independent Analysis (DIA) and Data Dependent Analysis (DDA) proteomics workflows and is compatible with variant call format (vcf) files from NGS workflows to enable personalized database searches. To assess quality of the resulting data, PAS includes various metrics like peptide/protein group intensity, protein sequence coverage, relative protein abundance distribution, peptide and protein groups. Visualizations including principal component analysis, hierarchical clustering, and heatmaps for which allow intuitive identification of experimental trends. To enable biological insights, differential expression analyses results are reported with interactive visualizations such as volcano plots, protein interaction maps, and protein-set enrichment. From data to insight, PAS provides an easy-to-use and efficient suite of tools to enable proteogenomic data analysis.

Integration of proteomics and genomics data require a variety of tools, many of which require command-line interfaces and operating system-specific specifications that can act as a barrier for researchers to adapt new data analysis tools. Here, we demonstrate the utility of PAS by analyzing 141 Proteograph plasma samples1 and performing a database search (tryptic; CID/HCD fragmentation; 25 ppm fragment and precursor tolerance; FDR threshold 0.01). 21,959 peptides and 2,499 protein groups were identified. This search was launched through the user interface requiring only three clicks, and in the background this search provisioned 142 servers and completed in approximately five and half hours. Together, these results show the utility of PAS for seamless and fast proteomic data analysis.

Reference: 1Blume, J. E. et al. Nature Communications, 2020

OmicsPipelines: a cloud-native desktop and web application for launching and managing Cromwell-based proteomics and genomics pipelines

Informatics and Computational Biology

POSTER 210

Mihir Samdarshi[1], **David Jimenez-Morales**[1], **Archana Raja**[1], **Steve Hershman**[1], **Matthew Wheeler**[1,2], **Euan Ashley**[1,2,3,4]

1 Division of Cardiovascular Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, CA
2 Stanford Cardiovascular Institute, Stanford University, Stanford, CA
3 Department of Genetics, Stanford University School of Medicine, Stanford, CA
4 Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA

The generation and analysis of large-scale multi-omics datasets are the new normal in biomedicine often requiring powerful computational resources. Due to the complexity of cloud environments, the availability of tools to enable the analysis of these datasets is still either very limited in scope or to a single cloud provider. Here we present an open-source application for more easily launching analysis pipelines of large-scale genomics and proteomics datasets in cloud environments. By providing a convenient user interface and default configurations without unnecessary bloat, OmicsPipelines attempt to minimize 1) the numerous verbose files which configure data processing pipelines, and 2) the complex behemoth that is cloud platforms. We illustrate that OmicsPipeline can easily be set up and run with limited knowledge of cloud computing.

The OmicsPipeline application consists of two modules: an installer and a dashboard. Both present an easy-to-use graphical user interface. The "installer" is a cross-platform desktop application that automatically creates all the necessary infrastructure components in a chosen cloud provider (Amazon Web Services or Google Cloud Platform), including a virtual machine, cloud data storage, dependencies, and other components required to orchestrate Cromwell jobs via a simple form with (customizable) default values. Once the application is installed, the "dashboard" allows the user to interactively create, run, and monitor Proteomics, Genomics, Epigenomics, and Transcriptomics workflows.

The Proteomics workflows allow the identification, quantification, and downstream analysis of MS-based proteomics data, including both isobaric chemical tandem mass-tag labeling (TMT/TMTpro) and label-free quantification, and analysis of global protein abundance, post-translational modifications. Remarkably, it currently supports two different proteomics tools, MS-GF+ and MaxQuant. The Genomics workflows allow the standardized analysis of genome sequencing, epigenomics (ATAC-seq, RRBS), and transcriptomics (RNA-seq) data. Once a specific omics/experiment workflow is selected, the user is presented with a form to provide the basic input files and parameters required to run the pipeline. Multiple users can potentially create and access various OmicsPipeline instances in a single account, facilitating collaborative work. The extensible OmicsPipeline application framework provides extensive support through documentation, tutorials, video-tutorials, and forums to request help.

(continue to page 62)

(continued from page 61)

In summary, we introduce OmicsPipelines, a comprehensive cloud-native application for the analysis of high-throughput multi-omics experiments, enabling the adoption of cloud technologies and their key features. By having available fast, scalable, and reproducible pipelines, we can enable the easier reuse of data and support testing or adoption of newer software tools, and the analysis of small and gigantic datasets.

miTAR: an interpretable deep learning–based approach for miRNA target prediction

Tongjun Gu[1,2], **Mingyi Xie**[3,4,5], **Xiwu Zhao**[8], **William Bradley Barbazuk**[1,5,6], **Ji-Hyun Lee**[2,7]

1 Bioinformatics, Interdisciplinary Center for Biotechnology Research, University of Florida, Gainesville, FL
2 Division of Quantitative Sciences, University of Florida, Gainesville, FL
3 University of Florida Health Cancer Center, University of Florida, Gainesville, FL
4 Department of Biochemistry and Molecular Biology, , University of Florida, Gainesville, FL
5 Department of Biology, University of Florida, Gainesville, FL
6 Genetics Institute, University of Florida, Gainesville, FL
7 Department of Biostatistics, University of Florida, Gainesville, FL
8 Department of Ophthalmology & Visual Sciences, University of Michigan, Ann Arbor, MI

microRNAs (miRNAs) are small RNAs that alter gene expression at the post-transcriptional level. They have been shown to play important roles in a wide range of biological processes. Many computational methods have been developed to predict targets of miRNAs in order to understand miRNAs' function. However, the majority of these methods depend on pre-defined features that require considerable efforts and resources to compute and often prove suboptimal at predicting miRNA targets.

We developed a novel hybrid deep learning-based approach that integrates convolutional neural networks (CNNs) and recurrent neural networks (RNNs). CNNs excel in learning spatial features and RNNs discern sequential features. Therefore, our approach has the advantages of learning both the intrinsic spatial and sequential features in miRNA:target. The inputs for our approach are raw sequences of miRNAs and genes that can be obtained effortlessly. Comparing with multiple previous methods on test datasets and independent datasets, we demonstrated that our approach gains unprecipitated better performance. In addition, we examined the contribution of a Max pooling layer in between the CNN and RNN on multiple datasets and showed that it improves the performance of all our models. Finally, a unified model was developed that is robust on fitting different input datasets (https://github.com/tjgu/miTAR).

Furthermore, we investigated the features learned by miTAR on two positive and negative datasets. We proved with a serial of in-silico mutagenesis analyses and correlation analyses that miTAR captures known features in miRNA:target, including the involvement of seed region and the free energy, as well as multiple novel features. Interestingly, the CNN and RNN layers of the models behave differently at capturing the free energy feature: the feature captured by the CNN layer units, but not the RNN layer units, is overlapped within and across feature maps. Although deep learning models are commonly thought "black-boxes", our discoveries support that the biological features in miRNA:target can be unveiled from deep learning models, which will enhance the understanding of the mechanism in miRNA's functional process and miRNA associated diseases.

DRAGEN advancements toward clinical genome applications: latest small variant calling accuracy gains in clinically relevant regions

Informatics and Computational Biology

POSTER 212

Severine Catreux[1], **Cooper Roddey**[1], **Lisa Murray**[1], **Fred Farrell**[1], **Mike Ruehle**[1], **Gavin Parnaby**[1], **Rami Mehio**[1]

1 Illumina Inc., Foster City, CA

Whole genome sequencing (WGS) continues to be increasingly adopted in clinical settings, for evaluating rare disorders but also for emerging applications such as preconception carrier screening, newborn screening or pharmacogenomics.

The Illumina DRAGEN Bio-IT Platform offers standardized informatics tools for the analysis of next-generation sequencing (NGS) data that can meet clinical genomics needs, by providing a fast, efficient and reliable way to analyze WGS data on a large scale, which is key when translating genetic variant information to medical diagnostics.

With the recent advances of high confidence call sets in extended regions of the genome which include clinically relevant genes , as well as new benchmarks developed for challenging medically relevant genes (CMRG), we show how the latest algorithm improvements in DRAGEN's mapper +variant caller make the platform ideally positioned to turn genomics finding into clinical decisions, with one simple but comprehensive sample-to-answer workflow.

Using the NIST v4.2.1 extended truth sets released for 7 subjects HG001-7, we show that the DRAGEN accuracy gains first presented in the PrecisionFDA Truth Challenge V2 with a 50% total error reduction using the graph genome , consistently hold across all 7 samples, highlighting that DRAGEN graph genome is not sample-specific and indeed extend to more samples and populations. The recently integrated machine learning methods also show consistent gains across all 7 samples, with ~30% FP reduction, that are additive to the gains from the graph.

We also show how DRAGEN graph genome is equipped to deal with high level of polymorphism in MHC and yields much improved variant calling accuracy in these regions compared to previous DRAGEN versions.

The combinations of these improvements alongside targeted callers for more complex genes, such as CYP2D6, SMN1/2, GBA, HLA-A-B, are also assessed against the CMRG truth set.

Via the combination of graph genome advances, integration of machine learning methods and development of targeted callers for more complex genes, we show that not only DRAGEN variant calling accuracy has significantly improved over previous versions, outperforming other read technologies specifically for INDELs, but it also provides an all-inclusive solution that can assess accurately clinical variants across the whole genome via a single test, in a scalable way that can process millions of genomes.

Spatial analysis of single-cell ligand-receptor expression elucidates tumor-immune interactions in lung cancer

Informatics and Computational Biology

POSTER 213

Zachary Reitz, **Patrick Danaher, Erin Piazza, Youngmi Kim, Mark Gregory, Sean Kim, Sarah Murphy, Nathan Schurman, Michael Leon, Emily Killingbeck, Bryce Kan, Michael Rhodes, Vikram Devgan, Joseph Beechem**

NanoString Technologies, Seattle WA

Tumors may escape or succumb to immune attack depending on how well they exploit a complex network of immunomodulatory ligand-receptor signaling. Studying these networks of interaction has proved challenging. Thus far, ligand-receptor interactions have been directly observable through low-plex imaging or indirectly estimable through high-plex single-cell RNA-sequencing. The latter of these techniques fails to represent cells within their proper spatial context. However, the advent of spatial transcriptomics technologies like the NanoString® CosMx™ has enabled us for the first time to infer ligand-receptor signaling across millions of physically interacting cells. Using CosMx data from non-small cell lung cancer (NSCLC) and a novel computational method, we search for evidence of coordinated ligand-receptor expression in neighboring tumor and immune cells.

Before examining tumor-immune interactions, we first detail the spatial relationships of tumor and immune cells. In an FFPE NSCLC, tumor cells fell into transcriptionally distinct clusters. These clusters comprised two spatially discrete layers within the tumor: a layer bordering the immune-rich microenvironment and an isolated interior layer. Notably, different T-cell populations neighbored tumor cells at varying frequencies. After quantifying these spatial relationships, we found that the immune adjacent tumor layer interacted far more frequently with CD8 memory and regulatory T-cells than naive and CD4 memory T-cells.

With the spatial map of tumor and T-cells in place, we turned to investigate ligand-receptor signaling between neighboring tumor cells and T-cells. To accomplish this goal, we devised a method for measuring spatial enrichment of pairwise ligand-receptor expression in neighboring cell pairs. We found several spatially enriched ligand-receptor pairs at the tumor-T cell interface. These included well-known signaling partners indicative of T-cell exhaustion and dysfunction, namely CTLA4/CD86 and PD-L1/PD-1. Additionally, pairs associated with cytotoxic T-cell function and complement signaling were also enriched. Most strikingly, immunomodulatory MHC class I to CD3 pairs were strongly upregulated at the interface. Using differential expression analysis, we found MHC Class I loci HLA-A, B, and C were all upregulated specifically in those tumor cells adjacent to T-cells.

By developing new methods for analyzing high plex spatial single-cell expression data, we can systematically interrogate context-dependent cell behavior within the TME. The data we present here is the first of many studies at this exciting new frontier of research.
FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Spatially regularized clustering algorithm for identifying cell types and tissue domains from spatial omics data

Informatics and Computational Biology

POSTER 214

Vipul Singhal[1*], **Nigel Chou**[1*], **Joseph Lee Jing Xian**[1,2], **Jinyue Liu**[1], **Reema Baskar**[1], **Erica Teo**[3], **Yun-Ching Chang**[3], **Lee Hwee Kuan**[4], **Kok Hao Chen**[1], **Shyam Prabhakar**[1]

1 Spatial and Single Cell Systems Domain, Genome Institute of Singapore, Singapore
2 National University of Singapore, Singapore
3 Veranome Biosystems, Mountain View, CA and Nucleos, Singapore
4 Bioinformatics Institute, Singapore
*Equal contribution

Spatially resolved transcriptomics technologies have revolutionized biological discovery by facilitating characterization of cell states within the spatial architecture of tissues. However, algorithm development has not kept pace. In fact, spatial omics data are still commonly analyzed using algorithms that ignore spatial information because they were originally designed for single-cell RNA sequencing data.

To address this challenge, we developed an algorithm for spatially regularized clustering of single cells. The algorithm, named Building Aggregates with a Neighbourhood Kernel and a Spatial Yardstick (BANKSY), incorporates information from a cell's immediate microenvironment into its own omics profile to generate an expanded, higher-dimensional feature set. BANKSY is general-purpose: it can be applied to data from diverse spatial technologies such as multiplexed fluorescence in-situ hybridization (multiplexed FISH), Slide-Seq, and multiplexed ion beam imaging by time of flight (MIBI-TOF). In benchmarking analyses on multiple datasets, we found that BANKSY improved cell clustering, crisply defined regional boundaries within tissues and uncovered a pair of closely related cell sub-types that inhabited distinct tissue microenvironments: 1) BANKSY accurately delineated cells within the granular layer, the Purkinje cell layer, and regions containing oligodendrocytes in a Slide-Seq dataset from mouse cerebellum that has proved to be challenging for conventional single cell clustering algorithms. 2) Applied to a multiplexed FISH dataset from the mouse hypothalamus, BANKSY identified two subtly different oligodendrocyte cell clusters originating from distinct hypothalamic regions. 3) In zone-finding mode, BANKSY accurately identified anatomical regions in a new multiplexed FISH dataset from mouse brain, as validated using the Allen Brain Atlas. 4) In a MIBI-TOF dataset from colon tumors, BANKSY automatically segmented tumor microenvironment zones previously defined using manual annotation. In summary, BANKSY is a versatile, general-purpose clustering tool that combines spatial location and molecular abundances to accurately detect cell types and tissue domains in spatial omics data.

Evaluation of low-pass whole genome sequencing as a cost-effective alternative to array-based imputation

Informatics and Computational Biology

POSTER 215

Peng Zhang, **Hua Ling, Justin Paschall, Beth Marosy, Kimberly Doheny**

Center for Inherited Disease Research (CIDR), Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD

As the cost of sequencing continues to drop, low-pass whole genome sequencing (lpWGS) has become a potential alternative for SNP arrays in genome-wide association studies. Compared to SNP arrays with pre-selected variant sites which often target specific populations, lpWGS is unbiased in terms of variant sites, limited only by the sites in a reference panel. The emergence of both new reference panels derived from large-scale diverse deep WGS data and more computationally efficient algorithms designed for imputation and phasing of lpWGS is now making lpWGS possible. Here we design and analyze an experiment to evaluate aspects of practical implementation including: library preparation methodology, impact across ancestry groups, DNA source, lpWGS coverage levels, imputation algorithm, imputation reference and compute resources.

We created a validation 'truth' dataset by deep sequencing 92 samples with diverse ancestry background (down sampled coverage to 28X). The lpWGS 'test' dataset (~ 1X) was generated using a cost-effective library prep (plexWell LP384 from seqWell) for the same set of 92 samples. For cost-comparison modeling, we mimicked NovaSeq6000 run conditions of 768 samples in a 2 x 150 bp S4 flowcell. We used two recent imputation methods designed for lpWGS, GLIMPSE and loimpute (from Gencove), to perform imputation based on the 1000 Genomes (1KG) deep sequencing reference panel (NYGC). The SNP array 'test' dataset was in silico array data derived by extracting the Illumina Global Screening Array (GSA, 654K variants) genotypes from the deep WGS validation dataset, and then performing imputation with the TOPMed reference panel.

We evaluate both sensitivity and concordance of each method by comparing all imputed sites in the validation dataset, limited to sites that pass some commonly used filters (R-square from TOPMed, INFO from GLIMPSE, max(GP) from GLIMPSE and loimpute). Our preliminary results from autosomal chromosomes for the 92 samples show that lpWGS followed by GLIMPSE imputation to 1KG achieves a comparable number of non-reference calls (median sensitivity 80.2% for lpWGS vs 81.5% for array). After applying filters, lpWGS has improved concordance over array, the average aggregated r2 (squared Pearson correlation between imputed and validation dosages) are 0.96 vs 0.94 for African, 0.93 vs 0.89 for Asian, and 0.95 vs 0.95 for European samples, and the gain is most extreme for low frequency variants (minor allele frequency 0.5%-10%). Our study demonstrates the feasibility of using lpWGS as a cost-effective and more population agnostic alternative for array-based approaches using publicly accessible resources.

Haplotypic diversity of centromeres and their evolution via long-read sequencing of >200 Japanese genomes

Informatics and Computational Biology

POSTER 216

Yuta Suzuki, **Shinichi Morishita**

The University of Tokyo, Tokyo, Japan

Our understanding of centromere sequence variation across human populations is limited by its extremely long nested repeat structures called higher-order repeats that are challenging to sequence. Recently, Here, we have reported an analysis of selected analyzed chromosomes (11, 17, and X) using long-read sequencing data for 36 individuals from diverse human populations, where including a Han Chinese trio and 21 Japanese.

We revealed substantial structural diversity with many previously unidentified variant higher-order repeats (HORs) specific to individuals characterizing rapid, haplotype-specific rapid evolution of human centromeric arrays, and conservation of , while frequent single-nucleotide variants are largely conserved.

While We found a characteristic pattern shared among prevalent variants in human and chimpanzee. these Our findings pave the way for studying sequence evolution in human and primate centromeres, we have scaled up the analysis to investigate how the current structure of haplotypic diversity of centromeres are formed within Japanese population, making the most of recent advances in base-level accuracy and throughput of long-read sequencing, and therefore utilizing an expanding resource of long-read sequencing data of >200 Japanese individuals..

We are going to present an analysis of variant HOR distribution within individual genomes which characterizes discrete haplotypes hidden in diploid data, as well as their degree of correlation to peri-centromeric SNPs defining centromere haplotypes (as known as cenhaps) to look into the evolution inside the centromeres.

Streamlining accessibility and computability of large-scale genomic datasets with the NHGRI genome data science Analysis, Visualization, and Informatics Lab-Space (AnVIL)

Informatics and Computational Biology

POSTER 217

Michael C. Schatz[1], **Anthony Philippakis**[2], **AnVIL Team***

1 Departments of Computer Science and Biology, Johns Hopkins University, Baltimore, MD
2 Broad Institute of MIT and Harvard, Cambridge, MA
*The full list of contributors is available at: https://anvilproject.org/about/team

The traditional model of genomic data sharing – centralized data warehouses such as dbGaP from which researchers download data to analyze locally – is increasingly unsustainable. Not only are transfer/download costs prohibitive, but this approach also leads to redundant siloed compute infrastructure and makes ensuring security and compliance of protected data highly problematic.

The NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-Space, or AnVIL (https://anvilproject.org/), inverts this model, providing a cloud environment for the analysis of large genomic and related datasets. By providing a unified environment for data management and compute, AnVIL eliminates the need for data movement, allows for active threat detection and monitoring, and provides elastic, shared computing resources that can be acquired by researchers as needed. AnVIL currently provides harmonized access to more than 275,000 genomes from several key NHGRI projects, such as as the CCDG (Centers for Common Disease Genomics), CMG (Centers for Mendelian Genomics), eMERGE (Electronic Medical Records and Genomics), and GTEx (Genotype-Tissue Expression Project), with many more on the near horizon.

The platform is built on a set of established components that have been used in a number of flagship scientific projects. The Terra platform provides a compute environment with secure data and analysis sharing capabilities. Dockstore provides standards based sharing of containerized tools and workflows. Jupyter, R/Bioconductor and Galaxy provide analysis environments for users at all skill levels to interactively explore and understand data with thousands of tools available. The Gen3 data commons framework provides data and metadata ingest, querying, and organization. Together, AnVIL provides a collaborative environment for creating, analyzing, and sharing data and analysis workflows for even the largest projects.

Long-term, the AnVIL will provide a unified platform for ingestion and organization for a multitude of current and future genomic and genome-related datasets. Importantly, it will ease the process of acquiring access to protected datasets for investigators and drastically reduce the burden of performing large- scale integrated analyses across many datasets to fully realize the potential of ongoing data production efforts.

Informatics and Computational Biology

POSTER 218

Toward understanding genomics cloud analysis costs

Enis Afgan[1], **Bridget Carr**[1], **Victor Wen**[1], **Peiyuan Xu**[1], **Keith Sudderman**[1], **Dannon Baker**[1], **Michael C. Schatz**[1]

1 Johns Hopkins University, Baltimore, MD

There is an increasing number of genomic datasets being deposited exclusively in the cloud. This includes commercial offerings, such as the Amazon Open Data program or the Google Cloud Life Sciences data, and a growing number of federally funded projects that provide access to high-value datasets alongside analysis applications, such as NHGRI AnVIL, NHLBI Biodata Catalyst, or the UK Biobank. The AnVIL alone houses nearly 300,000 whole genome datasets spanning nearly 4 Pb of data.

This trend is not reversing; modern datasets are large making them difficult to host or move. They also often require specialized environments that guarantee adequate security and compliance measures. As a result, an increasing number of data analyses will also move to the cloud. However, when moving the analysis into the cloud, researchers are faced with the challenge of understanding the associated costs. This applies at the level of a project where a consortium needs to estimate the analysis costs of hundreds or thousands of analyses in bulk. It also applies at the level of a PI or a graduate student where they need to estimate costs at the level of a single workflow (e.g., RNA-Seq) or a tool (e.g., assembly). Yet, cloud service providers offer no guidance on either of these estimates. Furthermore, these providers offer limited guardrails on capping spend. In turn, a novice researcher or a runaway script may unexpectedly exhaust a budget prematurely.

To help develop an understanding of the associated analysis costs, we are building a resource reference database and cost analysis for a variety of popular genomics tools and workflows. We have identified the popular tools and workflows from historic usage of one of the most used public servers for genomic data analysis, Galaxy US (usegalaxy.org), which includes many popular tools such as bowtie, bwa, samtools, STAR, and many others. However, this historic data is noisy and difficult to use to interpret the impact resource and tool properties have on the resource requirements and subsequent costs. Hence, we are benchmarking the identified tools and workflows in a systematic fashion to develop a database with controlled variables that can be used to infer costs. Here, we will present our findings as well as showcase an interactive dashboard where the results can be consumed by anyone, manually or programmatically, to obtain accurate estimates about resource requirements and estimated costs.

Informatics and Computational Biology

POSTER 219

Automated annotation of human centromeres

Andrey V. Bzikadze[1], **Olga Kunyavskaya**[2], **Tatiana Dvorkina**[2], **Ivan Alexandrov**[2], **Pavel A. Pevzner**[3]

1 Graduate Program in Bioinformatics and Systems Biology, University of California, San Diego, CA
2 Center for Algorithmic Biotechnology, Institute of Translational Biomedicine, Saint Petersburg State University, Saint Petersburg, Russia
3 Department of Computer Science and Engineering, University of California, San Diego, CA

Recent advances in long-read sequencing technologies led to rapid progress in centromere assembly. Last year, we presented centroFlye — the first automated tool for centromere assembly from long error prone reads (Bzikadze and Pevzner, Nature Biotechnology, 2020). The same year, we joined efforts with the Telomere-to-Telomere Consortium that, in May 2021, announced a landmark achievement — the first complete assembly of the whole human genome (Miga et al., Nature, 2020; Logsdon et al., Nature, 2021; Nurk et al., bioRxiv 2021). This progress opened a possibility to address the long-standing questions about the architecture of human centromeres.

Centromeres are tandem repeats that are formed by units repeated thousands of times with limited nucleotide-level variations but extensive variations in copy numbers in the human population. Each unit represents a tandem repeat formed by smaller repetitive building blocks thus forming a nested tandem repeat. Partitioning all blocks into n clusters of similar blocks defines n monomers. A canonical order of monomers (referred to as a higher-order repeat or HOR) is specific for each centromere and is defined as the ancestral order of monomers that has evolved into the complex organization of extant centromeres. The current view of the centromere evolution is summarized by the Centromere Evolution (CE) Postulate: each centromere has evolved from a single ancestral HOR formed by different monomers. Although the CE postulate is universally accepted, we are not aware of a rigorous proof of this postulate or an algorithm that, given an extant centromere, derives its HOR. We developed CentromereArchitect that addresses both the monomer inference and HOR inference as two separate problems (Dvorkina et al., Bioinformatics, 2021). However, more recent analysis revealed that, to generate a biologically adequate centromere annotation, the monomer and HOR inference should be viewed as two interconnected problems.

We present the HORmon algorithm that incorporates the monomer and HOR generation into a single pipeline and generates the first automated centromere annotation that is largely consistent with the CE Postulate and previous manual centromere annotations (Uralsky et al., 2019). Recognizing that HORs represent an important evolutionary concept, we show how HORmon can be used to automatically derive the currently known HORs.

Informatics and Computational Biology

POSTER 220

The spatial transcriptomic landscape of the healing intestine following damage

Sara M. Parigi[1], **Ludvig Larsson**[3], **Srustidhar Das**[1,2], **Ricardo O. Ramirez Flores**[4], **Annika Frede**[1,2], **Kumar P. Tripathi**[1,2], **Oscar E. Diaz**[1,2], **Katja Selin**[1,2], **Rodrigo A. Morales**[1,2], **Xinxin Luo**[1,2], **Gustavo Monasterio**[1,2], **Camilla Engblom**[5], **Nicola Gagliani**[1,2,6], **Julio Saez-Rodriguez**[4], **Joakim Lundeberg**[3], **Eduardo J. Villablanca**[1,2]

1 Division of Immunology and Allergy, Department of Medicine Solna, Karolinska Institute and University Hospital, Stockholm, Sweden
2 Center of Molecular Medicine, Stockholm, Sweden
3 Science for Life Laboratory, Department of Gene Technology, KTH Royal Institute of Technology, Stockholm, Sweden
4 Heidelberg University, Faculty of Medicine, and Heidelberg University Hospital, Institute for Computational Biomedicine, Bioquant, Heidelberg, Germany
5 Department of Cell and Molecular Biology, Karolinska Institute, Stockholm, Sweden
6 Department of Medicine and Department of General, Visceral and Thoracic Surgery, Hamburg, Germany

The intestinal barrier is a tissue that is specialized in controlling absorption or containment of luminal content. The intestine is known to be a highly compartmentalized organ, where segments have adapted to the heterogeneous exposure of microbes and nutrients. However, details about the compartmentalization of the colon are limited. To investigate the molecular and cellular circuitry of the colon, we used spatially resolved transcriptomics (SRT) technology. By folding the colonic tissue into a roll, we managed to fit tissue sections of the entire murine colon onto SRT glass slides, thus providing a map of the whole organ in one experiment. Then, to facilitate transcriptomic analyses along the proximal-distal axis of the colon, we developed a semi-automated computational unfolding procedure. Using data-driven analysis methods, we defined and characterized previously unappreciated compartments of the colonic epithelium.

Furthermore, we set out to investigate the principles of tissue repair of the intestinal barrier. For this purpose, we used colon tissue from a DSS-induced colitis model, and applied SRT during the recovery phase. We demonstrate dramatic global changes to the colon during this recovery phase and identified multiple transcriptional programs associated with mucosal healing. Overall, we provide a publicly available map of the transcriptomic landscape of the entire murine colon at steady state and during damage-induced tissue regeneration.

Informatics and Computational Biology

POSTER 221

BLR: a flexible pipeline for analysis of multiple linked-read techn

Pontus Höjer[1], **Tobias Frick**[1], **Humam Siga**[1], **Parham Posbourgh**[1], **Hooman Aghelpasand**[1], **Marcel Martin**[2], **Afshin Ahmadian**[1]

1 Royal Institute of Technology, Department of Gene Technology, Stockholm, Sweden
2 Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden

Several linked-read technologies have emerged in the last years. These use different approaches to co-barcode short subsections of long genomic DNA fragments. This facilitates the retrieval of long range information from short read sequencing. Linked reads thereby leverage the high base accuracy needed for small variants with the structural insights akin to long read sequencing. Despite the plurality of linked-read technologies, no common solution for analysis exists. LongRanger, a popular analysis pipeline, tailors to the now discontinued 10x platform for which it was developed. This makes integration of other technologies difficult and in some cases impossible due to technical differences.

We have developed the Barcode Linked Reads (BLR) pipeline for analysis of linked reads from the 10x Genomics, Droplet Barcode Sequencing (DBS), TELL-Seq and stLFR platforms. It takes raw reads from each of these platforms and generates phased alignments and small variant as well as large structural variant calls. The pipeline integrates all these technologies and accounts for varying barcoding strategies and library constructs. It uses several open-source tools including some from the growing collection developed specifically for linked reads.

To compare the pipeline to LongRanger we downloaded a recent 10x NA12878 dataset and run it though the BLR pipeline. To evaluate results we looked at the size of phase blocks which are continuous stretches of haplotype-resolved regions. We observed improvements in phase block N50 from 10.1 megabases (Mb) to 40.0 Mb for BLR with the longest being 121.6 Mb, spanning near the full long arm of chromosome 4. Switch errors were also considered to evaluate the accuracy of the phasing. Using a reference set of high confidence phased calls, switch errors were found to be at a low (<1%) level similar to LongRanger. Runs on other technologies have shown the same large phase blocks and low switch error rates indicating great phasing sensitivity and accuracy across platforms.

Informatics and Computational Biology

POSTER 222

A haplotype-resolved assembly of the major histocompatibility complex regions in the human genome

Bansho Masutani[1], **Shinichi Morishita**[1]

1 Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan

With the development of long-read sequencing, the problem of genome assembly is beginning to be solved, and the gapless assembly of the human genome by the T2T project would be a milestone. In addition, long-read technologies have also become a powerful tool for a more difficult problem: diploid assembly. That is, because the human genome is diploid, the true picture of 'genome' is two sequences for each chromosome. To solve this problem, mainstream methods first assembly the "unphased" contigs, call variants on them, phase found variants, and reassemble reads on each haplotype.

However, this method is not optimal for regions with large structural variations between homologous chromosomes. One such region is the MHC region, which is located on chromosome 6.

In this study, we are working to solve the diploid assembly problem of this region with many structural variations and single-nucleotide-polymorphisms. To this end, we have invented two ideas. The first is to locally assemble small regions -- chunks -- with 2000 bp length and determine their order in the genome. In this way, we can characterize the existence of structural variations among homologous chromosomes by the presence or absence of these chunks. The second is to distinguish differences between homologous chromosomes from these locally assembled chunks. In other words, we "phase" these assembled chunks by single-nucleotide-polymorphisms between homologous chromosomes. Existing methods usually call SNPs and phase them based on alignments in the phasing step. On the other hand, we perform phasing based on how the alignment to a locally assembled chunk changes when we introduce mutation to the chunk. In other words, if a chunk is a "father-derived" sequence, an editing operation that makes it closer to a "mother-derived" sequence will improve the alignment score of the "mother-derived" reads.

With these two novel methods, we have developed software to perform diploid assembly using long reads with errors of up to 15%. On the simulated dataset, we have validated that the software could assemble 1Mbp diploid genomes with 0.12% divergence and several structural variations. Experiments using haplotypes in hg38 have also yielded promising results. We plan to apply this method to multiple samples of Japanese to analyze the MHC region in the human population. Currently, scaling is a challenge as it takes about an hour to assemble a 4 Mbp region in diploids.

Informatics and Computational Biology

POSTER 223

pySCE: accelerated single-cell entropy scoring identifies known and novel tumor initiating cell markers of non-small cell lung cancer (NSCLC)

Robert P. Sebra[1,2,3,4], **Daniel M. Charytonowicz**[1], **Arpit Dave**[1], **Alex Tsankov**[1], **Kristin Beaumont**[1,2]

1 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY
2 Center for Advanced Genomics Technology, Icahn School of Medicine at Mount Sinai, New York, NY
3 Black Family Stem Cell Institute, New York, NY
4 Sema4, a Mount Sinai Venture, Stamford, CT

Tumor Initiating Cells (TICs) are small subpopulations of cancer cells believed to play a role in the development and progression of numerous cancers including lung, breast, and colon cancer. Despite consensus on the existence of TICs, existing markers used to define these enigmatic cells, particularly in lung cancers, are often inconsistently expressed and non-specific towards enriching a pure progenitor-like subpopulation. As current approaches towards the identification of TIC markers are largely driven by phenotypic characterization and fluorescence-based sorting methods, there is an urgent need for a higher-throughput, data-driven approach for determining potential CSC markers. In this work, we introduce and apply a GPU-accelerated computational entropy scoring algorithm, named pySCE, demonstrated to identify putative progenitor cell markers in both primary and metastatic lung cancer single cells, and validate our approach with an analysis of known differentiation dynamics in healthy lung tissue. Our approach robustly recapitulates expression dynamics associated with healthy lung homeostasis, and uncovers both known and novel marker genes associated with TICs in primary and metastatic lung cancer tissue, with both prognostic and therapeutic implications. Ultimately, we will demonstrate the potential for the pySCE analysis pipeline to be extended across other cancer types as a means of expanding the potential pool of TIC markers while offering an accelerated tool for transcriptional entropy scoring on increasingly large single cell datasets. Data presented will include pySCE benchmarking that shows a 250 to 1000-fold computational speed improvement using the GPU-driven engine to capture developmental, differentiation, stemness or other expression-based entropy scoring at scale. Further, the entropy specific single cell annotations and scoring within the healthy and disease lung datasets will be described at the gene level.

Medical
Omics

Ancestry inference from targeted NGS tests to enable precision medicine and improve racial/ethnic representation in clinical trials

Medical Omics

POSTER 300

Francisco M. De La Vega[1], **Brooke Rhead**[1], **Sean A. Irvine**[2]

1 Tempus Labs, Inc., Chicago, IL
2 Real Time Genomics, Ltd., Hamilton, New Zealand

There are well-established racial and ethnic disparities in cancer incidence and outcomes, in part due to structural, socioeconomic, environmental, and behavioral factors. However, some of these differences can be attributed to biological factors, such as the frequency of cancer mutations that vary by ancestry. It is well known that diversity in clinical trials is low, with Blacks and Hispanics consistently underrepresented compared to their cancer incidence. Further, race and ethnicity reporting in clinical trials occurs infrequently, and it is missing in up to 50% of patient medical records and genomic profiling tests. Moreover, self-reported race/ethnicity does not accurately reflect genetic ancestry, disproportionately affecting admixed patients. Representative participation in clinical trials would help minimize disparities in outcomes and enable the assessment of biological differences that may determine differential efficacy of drugs for oncology and other indications. Thus, the ascertainment of diversity in real-world genetic testing and clinical trial cohorts is needed. Rather than relying on self-reported race/ethnicity labels, ancestry can be inferred directly from sequencing data collected during tumor profiling and other tests. Ancestry is usually inferred from genome-wide data, either array or whole-genome sequencing (WGS), using unlinked random markers and clustering methods; but this approach is inappropriate for targeted next-generation sequencing (NGS) gene panels or even whole-exome sequencing (WES) data. Instead, we selected 654 ancestry informative markers (AIMs) overlapping the coding regions of 648 cancer genes targeted by the Tempus xT NGS assay. We implemented a supervised version of the ADMIXTURE algorithm using our AIMs to infer global ancestry proportions at the continental level (i.e., Africa, America, Europe, East Asia, and South Asia) from WGS, WES, and NGS xT variant calls. We validated our methods by comparing our results with labels of unadmixed donors from the 1,000 Genomes Project (1KGP), simulated admixture from the latter, and local ancestry inference previously derived with the RFmix software for admixed individuals from the 1KGP and the ICGC PCAWG project. We show that this method can infer ancestry and admixture proportions from targeted NGS testing data collected from research-consented patients where race/ethnicity was missing, and report concordance with self-described labels where available. Our results suggest that inferred ancestry can facilitate research on ancestry correlates with cancer mutations and outcomes data. Furthermore, ancestry inference from targeted NGS tests can be used to monitor the diversity in enrollment of oncology and other clinical trials to improve representation and ultimately patient outcomes.

Analysis of recombinant adeno-associated viral (rAAV) genomic integrity and purity using long read sequencing and automated library preparation

Medical Omics

POSTER 301

Kirsten Copren[1], **Elena Burashnikov**[1], **Ying Chen**[1], **James M. Wilson**[1], **Cheng-Chang Lee**[2], **Eduardo Cervantes**[2]

1 Gene Therapy Program, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA
2 Miroculus, San Francisco, CA

The University of Pennsylvania's Gene Therapy Program (GTP) utilizes recombinant adeno associated viral (rAAV) vectors in the pursuit of engineering vehicles to deliver therapeutic genetic material, such as a working gene, directly into a cell. For an IND (Investigational New Drug) enabling study, the FDA requires complete sequencing of the vector genome to ensure purity and integrity of the molecule. We currently have a rigorous, short-read sequencing assay in place to assess a variety of potential contaminants and to confirm the sequence, but short read sequencing is incapable of detecting larger structural variants and translocations, something long read sequencing can assess. We present the results of our collaboration with Miroculus, who optimized an automated protocol on its Miro Canvas instrument for the DNA Ligation Sequencing Kit (SQK-LSK110) from Oxford Nanopore Technologies. We compare the results of performing long read sequencing on ~35 rAAVs (a typical annual quantity) with pre-existing short read sequencing results to improve the level of detection for integrity and purity for our overall assay. Both methods allow for direct, amplification free sequencing of rAAV vectors with no additional PCR bias. Our short read sample to data workflow takes four days. The long read sample to data workflow takes three days. Our short read sequencing results find more single nucleotide variants, mainly in the ITR (Inverted Terminal Repeat) regions and the promoter, but no variants of concern. Occasionally, host DNA is found <1% of the sample. We achieved much higher coverage using short read sequencing. The long read sequencing library preparation with automation workflow using the Miro Canvas is simple and fast and will be a powerful tool for increased speed in our regulatory assay repertoire. This work is necessary to anticipate increased exigencies by regulatory bodies with the advancement of sequencing technology.

Quantitative spatial mapping of complex host response dynamics to SARS-CoV-2 infection

Medical Omics

POSTER 302

Silvia Groiss[1], **Daniela Pabst**[1], **Esther Föderl-Höbenreich**[1], **Melina Pabst**[1], **Kurt Zatloukal**[1], **Esther Föderl-Höbenreich**[1*]

1 Diagnostic & Research Institute of Pathology, Medical University of Graz, Graz, Austria
*Presenting author

Recurrent challenges to humans exemplified by the current COVID-19 pandemic highlight the importance of rapidly adapting scientific strategies to reach timely solutions. With the emergence of new SARS-CoV-2 variants, it has become increasingly important to understand the high degree of heterogeneity in host-pathogen interactions in response to the viral infection. Spatial transcriptomic approaches provide an excellent opportunity to study such heterogeneity. However, current methods are often marred by poor spatial resolution and sensitivity. Therefore, it is often challenging to get quantitative information on the location of single RNA transcript expression, moreover for the transcripts that are lowly expressed. To overcome these limitations, in this study, we resorted to a new spatial multiplexed single-cell mRNA in situ hybridization technology called Molecular Cartography™ (MC) developed by Resolve Biosciences. By implementing MC on various cell lines together with different SARS-CoV-2 variants, we have gained a thorough comparative understanding of host response dynamics. First, we demonstrate that MC detects single mRNA species at a specificity of > 99 % and that the sensitivity of mRNA detection is comparable to conventional single-molecule fluorescent in situ hybridization techniques. Furthermore, using SARS-CoV-2 infection models in cell lines of pulmonary, colorectal, and hepatocellular origin, we show that SARS-CoV-2 nucleocapsid protein transcripts prominently form a ring of expression surrounding the cell nucleus. Interestingly, we observed cell-to-cell variations in expression levels of the key viral entry factors ACE2, TMPRSS2 and FURIN. Our data also reveals virus-induced dysregulation of major primary and secondary antiviral genes. Finally, by spatially tracking antiviral response dynamics emanating from the primary infected cell , we elucidate differences in individual, temporal divergent infection sites and trace individual regulatory signatures in JAK/STAT pathways hijacked by SARS-CoV-2. Overall, Molecular CartographyTM allows to precisely detect and allocate multiple RNA species facilitating research on antiviral defence mechanisms that are essential to grasp and counteract SARS-CoV-2 pathologies.

Development and benchmarking of a SARS-CoV-2 Oxford Nanopore variation calling workflow

Medical Omics

POSTER 304

Chunlin Xiao, **Ryan Connor, Vadim Zalunin, Valerie Schneider, J. Rodney Brister**

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD

Rapid and accurate identification SARS-CoV-2 genetic variations in samples from infected patients is critical to the characterization and surveillance of the virus during the ongoing pandemic. Oxford Nanopore Technology (ONT) provides a rapid and economical sequencing platform, but has elevated error rates compared to other technologies. The suitability of ONT for use in SARS-CoV-2 genetic surveillance remains unclear and has yet to be comprehensively investigated. We have evaluated various strategies for read quality control, error correction, alignment, consensus generation, variation discovery, and variant refinement for ONT amplicon data, as well as whole genome sequencing data, and established a practical and efficient workflow for accurate identification of single nucleotide variations (SNVs) from noisy ONT sequencing data. We benchmarked the performance of SNV detection using ONT and Illumina sequencing data obtained from the same 76 patient samples and identified ONT-specific false positives. We then designed a workflow that produced comparable SNVs from reads sequenced by both ONT and Illumina. Finally, we applied this workflow to 74,949 ONT sequenced SARS-CoV-2 genomes available from the Sequence Read Archive (SRA). Of the small number of high frequency SNV positions observed across ONT sequenced SRA-CoV-2 genomes, many were shared with those detected in 156,000 Illumina sequenced SARS-CoV-2 genomes. Our findings suggest that with the suitable workflow, long nanopore sequencing reads could offer valuable information for accurate SNV identification and help establish linkage between SNVs within SARS-CoV-2 samples.

Barcode Enabled Antigen Mapping (BEAM) enables next-generation systems immunology analysis of the post–COVID-19 immune landscape

Medical Omics

POSTER 305

Michael Stubbington[1], **Bruce Adams**[1], **Daniel Reyes**[1], **Ariel Royall**[1], **Maengseok Song**[1], **Sean Marrache**[1], **Payam Shahi**[1], **Funien Tsai**[1], **Peter Finnegan**[1], **Thomas Vollbrecht**[1],**Tanmayee Khadilkar**[1], **David Jaffe**[1], **Ravi Ramenani**[1], **Wyatt McDonnell**[1], **Augusto Tentori**[1*]

1 10x Genomics, Pleasanton, CA
*Presenting author

Immune responses to disease are incredibly complex and the only way to fully understand such a response is to consider the behaviour of multiple different cell types across both the innate and adaptive branches of the immune system. Furthermore, we must perform multimodal measurements on large numbers of these cells to assess gene expression, protein expression and, in the case of T and B lymphocytes, antigen receptor sequences and specificities to a wide range of potential antigens.

We used Barcode Enabled Antigen Mapping (BEAM) and Immune Profiling technology to simultaneously profile all the above modalities at single cell resolution in hundreds of thousands of peripheral blood mononuclear cells (PBMCs) from a human donor following recovery from COVID-19. In addition to measurements of gene and protein expression, we generated full-length, paired sequences of the rearranged T- and B-cell receptors whilst also screening their specificity for a wide range of antigens from SARS-CoV-2 and other viral pathogens including cytomegalovirus, Epstein-Barr virus, and influenza.

These data provide unprecedented insights into the entirety of the immune landscape after recovery from acute viral disease. The scale and throughput of our experiments gave us high-resolution data from all cell types from the innate and adaptive immune systems.

We generated these data in less than one week from a single experiment and we were able to identify antigen-specific clones of both B and T lymphocytes, with the high cellular throughput enabling detection of rare clones that would not have been observed in smaller scale experiments. Furthermore, analysis of all PBMC cell types allowed us to place these antigen-specific clones within the overall transcriptional landscape of the immune system following recovery from viral infection.

Single cell, high throughput, and multimodal experiments such as these will underpin new systems immunology experiments and will continue to reveal the complex interplay between the components of the immune system. We envisage that these approaches will be extremely valuable in the analysis of the immune response to vaccination, infectious disease, cancer, allergy, autoimmune conditions, and ageing that can potentially lead to the development of novel diagnostic and therapeutic approaches.

Characterizing copy number variations using next- and third-generation sequencing and their association with plasma protein biomarkers

Medical Omics

POSTER 306

Daniel Schmitz[1], **Zhiwei Li**[2], **Nima Rafati**[3], **Åsa Johansson**[1]

1 Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Sweden
2 Twelve Bio ApS, Copenhagen, Denmark
3 National Bioinformatics Infrastructure Sweden, Department of Medical Biochemistry and Microbiology, Science for Life Laboratory, Uppsala University, Sweden

Introduction: Over the last years, genome-wide association studies (GWAS) have identified thousands of associations between single nucleotide polymorphisms (SNPs) and risk of diseases as well as with variation in disease-related traits. Structural variations (SVs), including copy number variations (CNVs), account for approximately 13% of genetic variability in humans. However, array-based SNP genotyping and CNV detection methods, such as array comparative genomic hybridization, lack the ability to detect novel signals or have low resolution compared to approaches based on next-generation sequencing (NGS).

Rational and aim: Recently, there has been an increasing interest in the role of CNVs in human development and diseases. To explore the role of CNVs in human disease pathology, we have identified CNVs and tested for association with plasma protein biomarker measurements, proteins that are well known or exploratory biomarkers for human diseases.

Methods: Whole genome sequencing was performed, in a Swedish cohort (N= 1021) using Illumina short-read technology (X-ten) to at least 30x per individual, and CNVs were called using CNVnator. A total of 438 plasma biomarkers were measured using Olink Proximity Extension Assay in 892 participants from the same cohort. We constructed a population-wide CNV matrix (CNV regions x individuals), assigning copy numbers to CNV regions according to CNVnator's calls. Including the 872 participants with both CNV and biomarker data available, we tested for association between the CNVs and biomarker levels, using linear regression analyses with age and sex as covariates. A set of five polymorphic CNVs were validated in 15 samples using Pacific Bioscience (PacBio) SMRT sequencing on a PacBio SEQUEL II system.

Results: A total of 243,987 polymorphic non-overlapping CNV loci were identified and included in the CNV-matrix. Significant associations were ($p<4.79\times⟦10⟧^{(-9)}$) were found between the copy numbers of 30 independent CNVs and the expression level of 17 protein biomarkers. Out of these CNVs, the breakpoints for two CNVs were validated by PacBio and two CNVs were identified to be clusters of many short repetitive elements. The fifth CNV could not be validated conclusively.

Conclusions: Our findings suggest that CNVs play an importation role in regulating the expression levels of protein biomarkers. These results provide insight into the possible effect of CNVs on human health and highlights the application of novel sequencing technologies for CNV detection. However, we also showed that short-read and long-read sequencing can provide non-comparable results, which suggests that long-read sequencing might be required to resolve complex SV events.

Spatially resolved analysis of the placenta tissue to study its transcriptomic landscape in preeclampsia

Medical Omics

POSTER 307

Nayanika Bhalla[1], **Lovisa Franzen**[1], **Annika Scheynius**[2], **Susanne Lager**[3], **Nikos Papadogiannakis**[4], **Stefan Hansson**[5], **Patrik Ståhl**[1]

1 Science for Life Laboratory, Division of Gene Technology, KTH Royal Institute of Technology, Stockholm, Sweden
2 Department of Clinical Science and Education, Karolinska Institute, Stockholm, Sweden
3 Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden
4 Department of Laboratory Medicine, Karolinska Institute, Stockholm, Sweden
5 Department of Obstetrics and Gynaecology, Lund University, Lund, Sweden

The placenta is a vital organ essential for foetal growth and development as it provides the developing foetus with a myriad of nutrients and growth factors. The molecular pathways and pathophysiology leading to placental disorders remain largely unknown and understudied due to the heterogenous nature of the tissue amongst different biopsies, even when the sampling location is well defined and similar. Previous single cell sequencing studies have only been able to identify markers for few of the placental cell types, but these efforts have not been able to interrogate the spatial relationship between the cells. For diverse tissue types, spatial transcriptomics is a technology that can provide genome-wide transcriptome profiling of tissue sections at a spatial resolution. Spatial gene expression studies in the placental tissue provided us with insights into the molecular basis and clinical significance of one such pregnancy disorder preeclampsia (PE), which affects 3-8% of pregnancies worldwide. Preeclampsia can prove to be fatal both for the mother as well as the foetus and its cause is still unknown. We optimised and devised a spatial transcriptomics protocol for placenta tissue and looked at gene expression patterns between 3 pairs of matched healthy and PE placenta tissue samples. We successfully managed to identify different clusters corresponding to different morphological patterns and cell types within the tissues, which aligns with the existing research in the field. We also identified potential differences between the case and control samples. This is a first transcriptomic study of its kind to be done on placenta tissue and the investigation of the molecular mechanism of PE through spatial transcriptomics can further help in identifying the molecular mechanism of the disease and the therapeutic outcomes early in the pregnancy. This study can also set a benchmark for further spatial research within placental omics.

High-plex spatial single-cell in situ analysis of lung SARS-CoV-2

Medical Omics

POSTER 308

Arutha Kulasinghe[1], **Chin Wee Tan**[2], **Vikram Devgan**[3], **Emily Killingbeck**[3], **Youngmi Kim**[3], **Mark Gregory**[3], **Michael Leon**[3], **James Monkman**[1], **Ning Liu**[2], **Dharmesh Bhuva**[2], **Kirsty Short**[4], **Fernando Guimaraes**[1], **Gabrielle Belz**[1], **Melissa J Davis**[2]

1 University of Queensland Diamantina Institute, The University of Queensland, Brisbane, Queensland, Australia
2 The Walter and Eliza Hall Institute of Medical Research, Parkville, Melbourne, Victoria, Australia
3 NanoString Technologies, Seattle WA
4 The University of Queensland, Brisbane, Queensland, Australia

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused a broad spectrum of disease, ranging from asymptomatic infections to acute respiratory distress syndrome (ARDS). COVID-19 is the clinical manifestation of SARS-CoV-2 infection. Little is known about the host tissue and cellular responses associated with COVID-19 infection, symptoms, and disease severity.

A better definition of the pulmonary host response to SARS-CoV-2 infection is needed to understand viral pathogenesis. Here, we use NanoString's Spatial Molecular Imager (SMI) technology to determine spatially resolved quantitative single-cell transcriptomic changes driven by SARS-CoV-2 infection. Rapid autopsy COVID-19 samples (n=19) were prepared on a tissue microarray and equivalent number of uninfected control patients were characterized with 1000 RNA targets at a single cell level. SMI enabled cell typing and mapping of complex cell populations while preserving spatial context and highlighted differential cell types distribution in the lungs of COVID-19 patients compared to non-infected controls. Collectively, these data demonstrate that single cell in situ analysis is a powerful tool to characterize cellular heterogenicity, and identify cell type and gene signatures within morphologically intact tissues, offering new insights into the pathogenesis of SARS-COV-2.

Spatially resolved, single-cell atlas of human pancreatic islets of Langerhans in type 2 diabetic and healthy controls

Medical Omics

POSTER 309

Emily Killingbeck[1], **Vikram Devgan**[1], **Michelle Pherson**[2], **Maureen Donlin**[2], **Sarah Murphy**[1], **Bryce Kan**[1], **Mark Gregory**[1], **Youngmi Kim**[1], **Gina Yosten**[2], **Sarah Warren**[1], **Grant Kolar**[2]

1 NanoString Technologies, Seattle, WA
2 Saint Louis University, St. Louis, MO

Diabetes mellitus is one of the leading causes of morbidity and mortality in the U.S., and rates of type 2 diabetes are rising. At the epicenter of the metabolic dysfunction associated with diabetes are the Islets of Langerhans, which encompass thousands of endocrine micro-organs scattered throughout the pancreas essential for the maintenance of normal glucose homeostasis. The majority of diabetes research has been conducted in rodent models, and few single cell transcriptomic studies on dissociated tissue have been performed on healthy or diseased human islets. Those that have revealed distinct subpopulations of alpha, beta, and delta cells, which are the major islet cell types. However, the anatomical bases of these subpopulations – and how they are altered in the setting of diabetes mellitus – remains unknown. Here we use a novel spatial molecular imaging (SMI) technology developed by NanoString's researchers to generate a spatially resolved single cell transcriptome of human pancreatic islet cells in both type 2 diabetes mellitus and high BMI, healthy controls. SMI enabled mapping of the subpopulations of beta, alpha, and delta cells in pancreatic islets in spatial context, and, crucially, provided resolution on the relative abundance and distribution of these subpopulations in healthy controls and their re-distribution in the context of type 2 diabetes. Furthermore, we characterized the abundance and distribution of immune cell populations that were also present in the tissue, providing insight into the ongoing immune response in the context of tissue damage. This spatially resolved, single cell atlas of human pancreatic islets gives unprecedented insight into the functioning of these endocrine micro-organs in the context of health, and provides valuable characterization of the molecular and cellular changes that are associated with tissue dysfunction and disease.

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

A high-throughput and high-content platform to screen for optimized cell therapy phenotypes

Medical Omics

POSTER 310

David DeTomaso, **Michelle Tan, Alan Tomusia, Natalie Bezman, Joe Choe, Aaron Cooper, John Gagnon, Brendan Galvin, Jeffrey Granja, Adam Litterman, Nishant Mehta, Jonah Pierce, Sean Poust, Luisa Silva, Grace Zheng, Nick Haining, Tarjei Mikkelsen**

Arsenal Biosciences, South San Francisco, CA

Chimeric antigen receptor (CAR) T cell therapy has emerged as a key new tool in the treatment of human cancers, achieving unprecedented patient outcomes in difficult to treat cancers such as acute lymphoblastic leukemia. To extend this treatment modality to new cancer types and improve its efficacy, a variety of enhancements have been proposed. Our ICT (Integrated Circuit T Cell) cells include genetic perturbations to alter cell state, receptor strategies to target multiple tumor antigens, engineered expression of cytokines and chemokines, and variations in the chimeric antigen receptor (CAR) binding or signaling domains.

To effectively survey this complex design space, we have developed a massively parallel, in vitro platform for novel cell therapy discovery. This highly automated platform enables parallel characterization of thousands of cellular designs with diverse readouts, including T cell proliferation, cytotoxicity, cytokine secretion, and flow cytometry. Whole transcriptome profiling is an additional, key readout enabled by a cost-effective, multiplexed 3' digital RNA-seq protocol. As a comprehensive, global assessment of cell states, this readout allows for an in-depth characterization of phenotypic differences between engineered T cell designs and provides insight into underlying mechanisms.

Here we describe the use of this platform to optimize ICT cell design with a focus on a transcriptional profiling pipeline and analysis. Over 10,000 variations of genetic perturbations, logic gate components, and CAR signaling domains were initially filtered using pooled screens with bulk and single-cell sequencing readouts. The resulting designs were evaluated using our automated, arrayed platform to identify individual components and their combinations with enhanced function. The ICT cell components identified using this approach promote superior potency, expansion, and persistence and induce resistance to tumor microenvironment-associated suppression. Our high-throughput, data-driven approach to cell design and optimization provides a scalable platform that greatly accelerates the development of safe and efficacious ICT cell therapies.

Spatially resolved detection of T cell receptor clonality elucidates spatial relationships between TCR expression, immune infiltration, and cancer-associated pathways

Medical Omics

POSTER 311

Michelle Kriner[1], **Katrina van Raay**[1], **Jason Reeves**[1], **Kit Fuhrman**[1], **Erin Piazza**[1], **Joseph Beechem**[1], **Margaret Hoang**[1*]

1 NanoString Technologies, Seattle, WA
*Presenting author

As T cells mature, genes encoding T cell receptor (TCR) segments are somatically recombined to generate a diverse repertoire of receptors specific to unique antigens. The resultant TCR diversity, and subsequent clonal expansion events, are critical in understanding the adaptive immune response to pathogens and cancers. Many methods have been developed to determine specific clonotypes and overall TCR diversity present in various tissues; however, nearly all fail to capture spatial orientation and arrangement of T cells engaging with their microenvironment. Here, we present an expanded TCR profiling panel for the GeoMx® Digital Spatial Profiler that can be combined with the GeoMx Cancer Transcriptome Atlas (CTA) or Human Whole Transcriptome Atlas (WTA), representing the first spatial assay for simultaneous quantification of all functional TCR constant, variable and joining segments in situ. We validated the performance of the TCR probe pool in inflamed tonsil and cell pellet arrays, demonstrating comparable sensitivity and specificity relative to orthogonal methods. We next used the GeoMx TCR spike-in panel to characterize intra- and inter-patient TCR heterogeneity in a cohort of 68 T cell lymphomas. T cell lymphomas are characterized by a dominant clone, corresponding to the tumor, and a population of potentially tumor-targeting T cells. Our results demonstrate the ability to link the spatial context of TCR segment expression in both malignant and non-cancerous T cells with the presence of other immune cells and cancer-associated pathways. Together, the combination of our TCR spike-in panel with the CTA or WTA illuminates T cell phenotypes, signaling pathways, population dynamics, and transcriptomic changes, yielding an unparalleled view of the T cell response in any context.

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Smoking modulates different secretory subpopulations expressing SARS-CoV-2 entry genes in the nasal and bronchial airways

Medical Omics

POSTER 312

Yuriy O. Aleksyev[1], **Ke Xu**[1], **Xingyi Shi**[1], **Chris Husted**[1], **Rui Hong**[1], **Yichen Wang**[1], **Boting Ning**[1], **Travis Sullivan**[2], **Kimberly M. Rieger-Christ**[2], **Fenghai Duan**[3], **Helga Marques**[3], **Adam C. Gower**[1], **Xiaohui Xiao**[1], **Hanqiao Liu**[1], **Gang Liu**[1], **Grant Duclos**[1], **Avrum Spira**[1], **Sarah A. Mazzilli**[1], **Ehab Billatos**[1], **Marc E. Lenburg**[1], **Joshua D. Campbell**[1], **Jennifer Beane**[1]

1 Boston University School of Medicine, Boston, MA
2 Lahey Hospital & Medical Center, Burlington, MA
3 Brown University School of Public Health, Providence, RI

Proposed risk factors for viral infection with severe acute respiratory syndrome coronavirus 2 SARS-CoV-2 include age, sex, chronic obstructive pulmonary disease, cancer, and cigarette smoking. To investigate whether these risk factors increase viral infection by modulation of the base line expression of Viral Entry (VE) proteins ACE2, TMPRSS2, and CTSL, we examined gene expression profiles of 796 nasal and 1,673 bronchial samples across four lung cancer screening cohorts containing individuals without COVID-19. Smoking was the only clinical factor reproducibly associated with the expression of any VE gene across cohorts. With smoking, ACE2 expression was significantly up-regulated in the bronchus but significantly down-regulated in the nose. Expression of individual VE genes were not correlated between paired nasal and bronchial samples from the same patients. Single-cell RNA-seq of nasal brushings revealed that an ACE2 gene module was detected in a variety of nasal secretory cells with the highest expression in the C15orf48+ secretory cells, while a TMPRSS2 gene module was most highly expressed in nasal keratinizing epithelial cells. In contrast, single-cell RNA-seq of bronchial brushings revealed that ACE2 and TMPRSS2 gene modules were most enriched in MUC5AC+ bronchial goblet cells. The CTSL gene module was highly expressed in immune populations of both nasal and bronchial brushings. Deconvolution of bulk RNA-seq showed that the proportion of MUC5AC+ goblet cells was increased in current smokers in both the nose and bronchus but proportions of nasal keratinizing epithelial cells, C15orf48+ secretory cells, and immune cells were not associated with smoking status. The complex association between VE gene expression and smoking in the nasal and bronchial epithelium revealed by our results may partially explain conflicting reports on the association between smoking and SARS-CoV-2 infection.

Novel assay for determining rAAV gene therapy vector breakpoints using modified bases

Medical Omics

POSTER 313

Donald E. Selby[1], **Terrence Hanscom**[1], **Luis Soares**[1], **Daniel R. Browne**[2], **John Walsh**[2], **Michael Weiand**[2], **Jonas Korlach**[2], **Jason B. Wright**[1], **John F. Thompson**[1]

1 Homology Medicines, Inc., Bedford MA
2 Pacific Biosciences of California, Inc. (PacBio), Menlo Park, CA

Recombinant adeno-associated virus (rAAV) is an important gene therapy vector due to its low immunogenicity coupled with being the only known eukaryotic virus capable of targeted integration. rAAV genomes are packaged in capsids as either positive or negative single stranded DNA (ssDNA). Upon capsid disruption, the ssDNAs rapidly hybridize to their complementary partners, forming heteroduplex double stranded DNA molecules. When standard NGS library preparation methods are used, length and sequence data for individual rAAV ssDNAs are lost, which is important information when assessing viral DNA quality for therapeutic use.

Here, we solve this challenge by developing a novel NGS library preparation method that takes advantage of polymerase gap-filling properties in conjunction with modified bases to create distinct DNA junctions at breakage points, which can occur in AAV vectors of all serotypes. The pre-existing ssDNA consists of unmodified bases while DNA beyond the break point includes modified bases denoting the sites where strand breaks occurred during ssDNA packaging into the viral capsid. Using the Sequel II system from Pacific Biosciences, the modified bases can be discerned from unmodified bases, enabling identification of the synthetic complementary strands at high resolution. We tested this method on a gene therapy research construct that was breaking during packaging. Using this assay, we were able to precisely locate the genomic region that was suffering from fractures and measure the true lengths of the original rAAV ssDNA molecules.

rAAV ssDNA breakpoint data provides us with important information into how to design therapeutic vectors as well as more precise QC data. Partially filled capsids have been an ongoing FDA concern and the technology we have developed provides information that standard methods do not. This novel use of modified bases for the purpose of localizing DNA breaks can enable better vector designs and quality metrics of AAV production, potentially resulting in higher quality gene therapy and editing vectors.

Molecular CartographyTM analysis of 'healthy' and 'diseased' gingival tissues

Medical Omics

POSTER 314

Moritz Kebschull[1], **Annika Kröger**[1*], **Paul Weston**[1], **Devan Raindi**[1]

1 University of Birmingham, Birmingham, United Kingdom
*Presenting author

Periodontitis is the most common non-communicable disease in humans, caused by a destructive interaction between a dysbiotic biofilm and a dysregulated host immune-inflammatory response. Periodontitis leads to tooth loss and is independently associated with pre-mature mortality. Oral diseases (mostly periodontitis) are globally responsible for more years lost to disability than any other human disease. Importantly, the detrimental effects – both in the oral cavity and systemically - are especially critical in patients with rapidly progressing variants of the disease. The pathophysiological mechanisms underlying rapid progression are, however, only incompletely understood.

Our group has in the past, using the largest cohort in the field, assessed the transcriptomes of 'healthy' and 'diseased' gingival biopsies to identify critical signatures of disease progression. Our work has led to a change of the world-wide classification of periodontal diseases. Still, this body of work represented only 'mixed bags of tissue' and thus could not or only incompletely account for signatures of individual cell populations in specific layers of the gingival tissues.

Thus, to identify the spatial transcriptomic signatures that (i) define the transition from 'healthy' to 'diseased' status, (ii) help explain the potential differences of grade B ('moderate' progression) to grade C ('fast progression') periodontitis, and (iii) can be linked to specific bacterial-host interactions, we have assessed a sample of 'healthy' and paired 'diseased' gingival tissue samples from systemically healthy non-smokers using the Resolve Biosciences' Molecular Cartography™ technology, coupled to a meta-genomic sequencing of the associated subgingival biofilms using Illumina technology. Specifically, healthy non-smokers contributed a 'diseased' gingival papilla (with bleeding-on-probing, probing depth ⩾4mm, and clinical attachment loss ⩾3mm), and a 'healthy' papilla (no bleeding-on-probing, probing depth ⩽4mm, and clinical attachment loss ⩽4mm). Data were assessed for specific differences between disease groups using mixed model regressions. Clusters of characteristic signatures were identified using mixture model-based clustering whilst accounting for covariates.

Our analyses revealed that specific transcriptomic signatures in a spatial context exist that help differentiate 'healthy' and 'diseased' status, as well as different progression rates. These signatures can be contributed to defined cell populations and locations within the tissues. They are driven by specific microbial complexes and could help to improve early diagnostics of progressive disease, as well as inform targeted therapeutic approaches.

Characterization of CNBP expansions in myotonic dystrophy type 2 patients using nanopore Cas9-targeted sequencing

Medical Omics

POSTER 315

Rossato M.[1], **Centofanti F.**[2], **Alfano M.**[1], **Visconti V.V.**[2], **Maestri S.**[1], **D'Apice M.R.**[2], **Novelli G.**[2], **Botta A.**[1], **Delledonne M.**[1]

1 Department of Biotechnology, University of Verona, Italy
2 Department of Biomedicine and Prevention, University of Rome Tor Vergata, Rome, Italy

Myotonic dystrophy type 2 (DM2) is a multisystemic disorder caused by a (CCTG)n in intron 1 of the CNBP gene. The CCTG repeat tract is part of a complex (TG)v(TCTG)w(CCTG)x(NCTG)y(CCTG)z motif generally interrupted in CNBP healthy range alleles. In DM2 patients, the interval of CCTG repeat units is extremely wide, ranging from 75 to over 11.000 units. Molecular diagnosis of DM2 is challenging because of the CCTG expansion length and the high level of somatic mosaicism, which complicate genotypes-phenotypes correlations in patients.

A pilot group of DM2 patients (n=9) were sequenced with a target enrichment approach based on the Crispr-Cas9 system coupled to Nanopore long-read sequencing. Sequencing results showed a similar pattern at the 5' end for all the samples consisting of "TG", "TCTG" and "CCTG" blocks. Interestingly, only 2 DM2 samples showed the expected pure "CCTG" pattern, while the other 7 DM2 samples showed at the 3'end a mixed pattern consisting of "TCTG" and "CCTG" blocks, never described so far in DM2 patients. TP-PCR sequencing confirmed the presence of (TCTG)n repetitions in patients carrying the "TCTG" block. Southern blot analysis confirmed the capability of the method to sequence across large repeat expansions up to >40 kb in length, otherwise not feasible with traditional methods or other long-read sequencing approaches. In addition, the bioinformatic pipeline developed allows to recognize and precisely count the repeat pattern within the CNBP gene and to unravel atypical motif un-identified by previous studies.

In conclusion, our preliminary data open new exciting scenarios on DM2 disease. The discovery and the characterization of non-CCTG sequences within CNBP expanded alleles will allow more accurate genotype-phenotype correlations and permit a better stratification of patient in clinical trials.

Identifying causative structural variants for rare Mendelian disease directly from secondary variant calling pipelines with Fabric Genomics' GEM algorithm

Medical Omics

POSTER 316

Erwin Frise[1], **Barry More**[2,3], **Guying Liu**[1], **Lucia Guidugli**[4], **Stephen F. Kingsmore**[4], **Mark Yandell**[2,3], **Martin G. Reese**[1]

1 Fabric Genomics, Oakland, CA
2 Department of Human Genetics, University of Utah, Salt Lake City, UT
3 Utah Center for Genetic Discovery, University of Utah, Salt Lake City, UT
4 Rady Children's Institute for Genomic Medicine, San Diego CA

Structural variants (SVs) are the root cause for over 10% of Mendelian diseases and it is therefore essential to include them for clinical evaluation. Major barriers to using SVs in the diagnostic process are the low sensitivity and high false positive rate of current SV callers. Moreover, most interpretation pipelines use several SV callers, with the hope that consensus call sets prove more informative, but in our experience, this only further complicates the process. To reduce the SV number to a more informative set, current best practice involves substantial filtering, often biased towards a consensus from multiple callers. Filtering is poorly standardized, further complicating clinical evaluation and the development of cross-institutional protocols. Here we will demonstrate significant benefits for probabilistic joint optimization of evaluation, filtering, and prioritization of SVs. GEM is an AI-based Electronic Clinical Decision Support System (eCDSS) for scalable diagnosis of rare genetic disease by WES or WGS (Genome Medicine, in press). GEM incorporates OMIM, gnomAD, HPO and ClinVar knowledge, while automatically controlling for confounding factors, such as sex, consanguinity, and ancestry. GEM is compatible with Clinical Natural Language Processing (CNLP) derived phenotypes and operates with equal accuracy on singleton and trio genomic datasets. We have added a new algorithmic approach to GEM to support arbitrary numbers of secondary SV callers on an "as is", meaning that no prefiltering or heuristic consensus building are required. We tested this new GEM functionality using retrospective, solved WGS cases of seriously ill children from the Rady Children's Hospital (RCH) NICU. Briefly, FASTQ files were processed with DRAGEN, typically resulting in approximately 7,000 SV calls, and then imported into GEM. We also tested GEM with two RCH cases with SVs from 4 different SV callers. Our results show that despite large numbers of SVs, nearly 100% accuracy is obtainable using these real cases, with the causative SV generally ranked first among fewer than 10 candidate SVs, regardless of the number of SV callers used and number of SVs considered. This holds true even when no SV calls are provided, as GEM can reliably infer the presence of SVs directly from SNV/Indel VCF files. Joint optimization vastly simplifies SV processing and provides a superior and integrated clinical interpretation combined SV and small variants datasets.

Other Biology Omics

Molecular CartographyTM studies of how chronic pain and opioid treatment transform RNA expression patterns in the mammalian amygdala

Other Biology
Omics

POSTER **400**

Grégory Scherer[2], **Daniel Berg**[1], **Yuxi Ke**[1], **Xiqian Jiang**[1], **Mark Schnitzer**[1]

1 Stanford University, Stanford, CA
2 The University of North Carolina at Chapel Hill, Chapel Hill, NC

Neural activity in the amygdala has a crucial role in shaping the unpleasantness of painful experiences. For example, we recently found that a subset of neurons in the basolateral subdivision of the amygdala (BLA) respond selectively to painful stimuli, indicating the causal role of these cells in the experience of pain. Chemogenetic inhibition of these BLA nociceptive neurons in mice in a chronic neuropathic pain state markedly reduced affective-motivational pain-related behaviors (Corder et al., Science, 2019). Notably, several types of amygdalar neurons express opioid receptors, raising the possibility that modifications to neural signaling in the amygdala may be a primary means by which opioid treatment alleviates the experience of pain. Similarly, alterations in amygdala neural activity may contribute to the unpleasantness that accompanies withdrawal from opioid treatment. Given the joint epidemics of both chronic pain and opioid addiction, achieving a better understanding of how both pain and opioids affect the brain is an important step toward developing non-addictive methods for treating pain.

Thus, to identify molecular events that are evoked in amygdalar neurons by the experience of pain, opioid treatment, and withdrawal from opioid treatment, we are applying Molecular Cartography™ (Resolve Bioscences) to spatially localize single RNA molecules at sub-cellular resolution, using up to 100 different RNA probes with individual tissue slices. The set of genes that we are targeting with molecular cartography includes a set of marker genes that will allow us to differentiate the distinctive types of excitatory and inhibitory neurons in the amygdala, a set of genes that are indicative of neural plasticity (such as immediate early genes), as well as genes encoding opioid peptide precursors, opioid receptors and other G protein-coupled receptors. By mapping the spatially resolved patterns in which these genes are organized in the amygdala of control mice and mice that have undergone either a persistent painful experience, morphine exposure, morphine treatment after a painful experience, or withdrawal from morphine exposure, we aim to uncover fundamental molecular aspects of how the brain responds to these extraordinarily salient experiences.

Spatial and single-cell profiling of cerebral organoids reveal the cartography of gene expression during human cortical development

Other Biology
Omics

POSTER **401**

Benedikt Nilges[2], **Ilaria Chiaradia**[1], **Madeline A. Lancaster**[1], **Ivan Imaz-Rosshandler**[1], **Nachiket Kashikar**[2], **Richa Das**[2]

1 MRC Laboratory of Molecular Biology, Cambridge Biomedical Campus, Cambridge, United Kingdom
2 Resolve Biosciences, Monheim am Rhein, Germany

Cerebral organoids lay at the interface of cell biology and tissue biology. Indeed, the highly organised cytoarchitecture of proliferative and neuronal layers mimics faithfully the developing cortical tissue. This tissue context is an important part of cortical development. Widely used single-cell RNAseq techniques necessarily lose this spatial information due to dissociation of the tissue. Our work aims to combine single cell RNAseq data with spatial coordinates (Molecular CartographyTM) of organoid cell types, resulting in a cartography of cerebral organoid gene expression. We demonstrate how the two techniques can be combined and complement each other. This leads to the identification of several progenitor subpopulations in the proliferative region of the organoid along with neuronal subtypes resembling upper and deep layer cortical neurons. Such a degree of spatial resolution is commonly achieved through immunohistochemistry (IHC) techniques, looking at the protein localisation. Molecular CartographyTM from Resolve Biosciences offers a complementary approach but provides a much higher throughput readout of transcript levels along with protein expression. Indeed, the combination of the two approaches enables us to recapitulate the neurogenic waves happening in cerebral organoids. Neurogenesis in vivo and in vitro follows a timely and spatially controlled program. Deep layer (DL) neurons are generated first from the proliferative layers, followed by upper layer (UL) neurons in the so-called inside out fashion. In mature organoids, IHC reveals UL neuronal markers expressed throughout the cortical plate not overlapping with DL markers. However, when looking at transcript localisation for UL markers CUX1 and SATB2, they are highly expressed in progenitors of cerebral organoids, demonstrating the switch from DL to UL neurogenesis. Detection of such precise temporal resolution is only achievable by analysing gene expression at the transcript level. Taken together, our results suggest that transcriptome studies benefit from the integration with spatial information when it comes to investigating the identity of a highly complex tissue such as developing human brain tissue.

Other Biology
Omics

POSTER 402

Experience with whole-genome long-read sequencing in the setting of previously banked samples

Erin LaRoche, **Ally Day, Stacy-Ann Mano, Tera Bowers, Kiran Garimella, Katie Larkin, Stacey Gabriel**

Broad Institute of MIT and Harvard, Cambridge, MA

Human whole genome sequencing via long read sequencing platforms is an increasingly useful tool in gene discovery programs where short read data has failed to uncover potentially causative variation. While long read sequencing has a high degree of utility in detecting structural, complex types of DNA variation it tends to carry the requirement of very high quality and quantity of DNA to produce the most useful data. This can present a challenge in human disease studies where it is often the case that DNA samples from well-characterized cohorts have been collected over many years and banked for later use.

We have undertaken a program to generate whole genome long read sequence data on ~400 previously banked samples (collected 5-15 years ago) from cohorts assembled for structural birth defects. The samples have short read data available but the underlying clues to genetic etiology are still incomplete.

We present technical details of our process to maximize the success of these samples on the PacBio Sequel IIe platform. Two hundred samples were chosen for continuous long read (CLR) sequencing and 200 for high fidelity (HiFi) sequencing. Prior to library construction, sample quality was assessed to determine suitability for the standard library construction protocol. On the first pass, we observed a high QC failure rate (>96%) using standard input requirements for purity, concentration and fragment size. For those samples which failed to meet these standard input requirements, we grouped batches with similar quality metrics and tailored our library prep approach to each group in order to maximize success for every sample.

While this study is still in progress, our results show that with the new criteria and modified library prep, as many as 88% of samples are considered suitable for downstream sequencing. As of this submission, 113 samples of 179 attempted in sequencing met deliverable on first attempt, with another 66 requiring some top off. This study will allow us to correlate sequence coverage results with the incoming QC metrics for banked samples.

Other Biology
Omics

POSTER 403

Spatial single-cell characterization of SIV reservoirs in lymphoid tissues and B cell follicles in rhesus macaques

Mike Nekorchuk[1*], **Sean Kim**[2], **Youngmi Kim**[2], **Chi N. Chan**[1], **Daniel J. Newhouse**[3], **Michael J. Gale**[3], **Sarah Murphy**[2], **Mark Gregory**[2], **Sarah Warren**[2], **Jacob D. Estes**[1]

1 Vaccine and Gene Therapy Institute, Oregon Health and Science University, Beaverton, OR
2 NanoString® Technologies, Seattle, WA
3 Department of Immunology, University of Washington, Seattle, WA
*Presenting Author

HIV cannot be eradicated by antiretroviral therapies (ART) alone. Although lifelong suppression of HIV replication with ART seems possible, side effects, resistance and cost contribute to the necessity of finding an 'HIV Cure'. The major obstacle to eradicating HIV is the persistence of viral reservoirs despite extended ART, and can give rise to recrudescent infection when ART is stopped. An understanding of complex spatial host-viral interactions in situ that retains important contextual insight into the cellular immune neighborhoods and inflammatory landscapes is needed to define mechanisms driving HIV persistence.

Spatial profiling of lymph nodes from chronically SIV infected nonhuman primates (NHP; rhesus macaques) 8 weeks post infection with SIVmac239M was performed on FFPE tissue sections using NanoString® GeoMx® and Spatial Molecular Imager (SMI) platforms. GeoMx relies upon hybridization of affinity reagents (antibodies or RNA-ISH probes) to the target in a slide mounted tissue section. Indexing oligos are cleaved from the affinity reagents via UV light and quantified by NGS. SMI uses in-situ imaging technology based on nucleic acid hybridization cycles of fluorescent molecular barcodes to measure biological targets on intact sample. In this study, four FFPE lymph nodes from SIV+ NHPs were characterized with 78 proteins or 1563 RNA targets at a multicellular level with GeoMx, and with 1000 RNA targets (including viral RNAs) at a single cell level with SMI.

GeoMx profiling of 4 SIV+ lymph nodes demonstrated that Regions of Interest (ROIs) with active viral infections had significantly altered transcriptional programs when compared to uninfected ROIs. As a result, a number of different pathways in the infected cell neighborhoods were altered including IL-4, IL-17, and HLA signaling, interferon/inflammation, and B cell development. The GeoMx spatial analysis allowed the discovery of >200 genes with differential expression that would not have been observed with pan-tissue or single-cell RNA sequencing approaches. SMI profiling of serial sections from the same samples augmented the GeoMx observations and provided further elucidation of the transcriptional profile but at single cell resolution.

Spatial profiling of the samples provides insight beyond what is available from bulk analyses and shows the role of viral reservoirs in the context of the cellular and immune neighborhoods. These results improve our understanding of NHPs as a model system for HIV and may lead to potential treatment options to eliminate the entirety of the viral reservoir in the infected individual.

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

A simple approach to selectively remove abundant unwanted RNAs and improve the sensitivity of transcript detection in any species

Other Biology
Omics

POSTER 404

Lynne Apone[1], **Deyra N. Rodríguez**[1], **Bradley W. Langhorst**[1], **Kaylinnette Pinet**[1], **Gautam Naishadham**[1], **Jian Sun**[1], **Keerthana Krishnan**[1], **Fiona J. Stewart**[1], **Eileen T. Dimalanta**[1]

1 New England Biolabs, Ipswich, MA

The presence of highly abundant transcripts, with minimal biological interest in total RNA samples, creates a challenge to whole-transcriptome sequencing. These highly expressed transcripts dominate readouts, obscuring the detection of more informative lower abundance transcripts. Here, we present a simple, customizable approach to enrich for RNAs of interest by eliminating unwanted RNAs. This method is based on hybridization of probes to the targeted RNA and subsequent enzymatic degradation of the selected RNAs. The probe sequences confer RNA removal specificity and can be designed to deplete unwanted RNA from any organism.

We developed a simple, user-friendly web tool that enables custom depletion of any RNAs of interest without the aid of a bioinformatician. We used this web tool and depletion method to remove rRNA from total RNA of various species, including the pig Sus domesticus, mosquito Aedes aegypti and the archaea Thermococcus kodakarensis. Additionally, we used this approach to target coding RNAs in human total RNA, and supplemented an existing anti-rRNA probe set for depletion of both rRNA and the selected coding RNAs.

Using strand-specific RNA sequencing we measured depletion efficiency and transcript expression. We achieved high depletion efficiency (up to 99%) for all targeted RNAs across different species, while maintaining transcript abundance of non-targeted RNA. This translated into an enrichment of RNAs of interest and an increased depth of sequencing coverage.

The method described here is a simple and reliable solution to improve sensitivity in RNA-Seq studies. The design tool offers flexibility and control over the probe design process facilitating a customized and economical RNA sequencing experience. Importantly, the depletion and library construction method are amenable to high throughput sample preparation and robotic automation.

Investigating molecular correlates of cognitive impairment in ALS-FTD using spatial transcriptomics

Other Biology
Omics

POSTER 405

Joana Petrescu

Columbia University, New York, NY

Amyotrophic lateral sclerosis and frontotemporal dementia (ALS-FTD) is a neurodegenerative disorder with clinical presentations ranging from progressive paralysis to cognitive impairment. Approximately 15% of ALS-FTD patients initially presenting with motor symptoms also receive a diagnosis of dementia, but a majority of ALS-FTD patients demonstrate some level of cognitive impairment over the course of disease. Identifying molecular pathways that contribute to the development of cognitive deficits in ALS-FTD has thus far been limited by the quality of clinical information and postmortem tissue preservation as well as available technologies. We make use of a cohort of ALS-FTD patients with detailed cognitive profiling and high quality postmortem tissue preservation. Using novel highly multiplexed protein imaging and spatially resolved transcriptomics methods, we are able to quantify cell type composition and gene expression in situ in postmortem tissue sections from our cohort. These experimental advances enable the study of molecular mechanisms that are associated with cognitive dysfunction in ALS-FTD.

Other Biology
Omics

POSTER 406

Disease-associated cell states in atherosclerosis defined by spatial transcriptomics (Molecular CartographyTM)

Minna U Kaikkonen[1], **Tiit Örd**[1], **Valtteri Nurminen**[1*], **Aarthi Ravindran**[1], **Lari Holappa**[1], **Lea Mikkola**[2], **Tapio Lönnberg**[2], **Eloi Schmauch**[1,3,4]

1 A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio, Finland
2 Turku Bioscience Centre, University of Turku and Åbo Akademi University, Turku, Finland
3 Massachusetts Institute of Technology, Computer Science and Artificial Intelligence Laboratory, Cambridge, MA
4 Broad Institute of MIT and Harvard, Cambridge, MA
*Presenting author

Atherosclerotic coronary artery disease is one of the leading causes of mortality in the developed world. Atherosclerotic lesion development encompasses macrophage adhesion on the luminal surface, endothelial dysfunction, and smooth muscle cell dedifferentiation into other cell types. However, how these disease states situate in the lesions or interact with each other, remains unknown. Here we used the Molecular Cartography™ platform by Resolve Biosciences to confirm lesional cell states identified by single cell RNA-Seq to interrogate their location in mouse thoracic aorta during atherosclerosis progression. Furthermore, we identify specific subsets of genes that display complex spatial expression in single or many lesional cell states and map cell-to-cell interactions in situ. Our results demonstrate that the integration of spatial information with single-cell gene expression data is essential for deeper understanding of vascular biology and atherosclerosis associated cell states.

Other Biology
Omics

POSTER 407

Detection of pseudouridine modifications in human mRNAs using direct, long-read sequencing

Sepideh Tavakoli[1], **Mohammad Nabizadehmashhadtoroghi**[2], **Amr Makhamreh**[1], **Howard Gamper**[4], **Neda K. Rezapour**[3], **Ya-Ming Hou**[4], **Meni Wanunu**[3], **Sara H. Rouhanifard**[1]

1 Department of Bioengineering, Northeastern University, Boston, MA
2 Department of Mechanical Engineering, Northeastern University, Boston, MA
3 Department of Physics, Northeastern University, Boston, MA
4 Department of Biochemistry and Molecular Biology, Thomas Jefferson University, Philadelphia, PA

Pseudouridine is one of the most abundant and dynamic RNA modifications in mammalian cells and plays an important role in post transcriptional gene regulation. However, we still don't know the precise sequence coordinates of all pseudouridylated sites or the exact biological function. Previous studies have identified putative sites of modification using pseudouridine-specific chemical labeling and short-read sequencing; however, these methods are limited due to incomplete labeling and necessary amplification of the target RNAs that can lead to false positives and negatives. We have developed a direct-RNA nanopore sequencing pipeline and algorithm to independently validate previously identified pseudouridylated sites and generate a validated set of pseudouridylated targets for the field. Using our validated method, we have detected over 4000 putative unknown pseudouridylated sites in the HeLa transcriptome, many of which are enriched in genes that encode RNA binding proteins. Using this data, we model two types of hyper-modification: 1. Single pseudouridylated sites that are frequently modified and 2. Individual mRNAs with more than one site of pseudouridylation. This pipeline will help to propel epitranscriptomics field forward and enhance our understanding of the biological role of RNA modifications in mammalian mRNAs.

Other Biology
Omics

POSTER 408

Spatial omics analysis of the cellular immune landscape and their infiltration routes in a minimally invasive mouse model of myocardial infarction (MI)

Denis Schapiro[1,2], **Florian Wünnemann**[1,2], **Florian Sicklinger**[3,4], **Clara L. Heine**[3,4], **Florian Leuschner**[3,4]

1 Institute for Computational Biomedicine, Faculty of Medicine, University Hospital Heidelberg and Heidelberg University, Heidelberg, Germany
2 Institute of Pathology, University Hospital Heidelberg, Heidelberg, Germany
3 Department of Cardiology, University Hospital Heidelberg, Heidelberg, Germany
4 German Centre for Cardiovascular Research (DZHK), Partner Site Heidelberg, Heidelberg, Germany

Myocardial infarction is one of the most frequent manifestations of cardiovascular disease across Europe, accounting for about 20% of all deaths. Recent studies highlight the importance of tissue inflammation and the involvement of the immune system in the pathophysiology and subsequent healing of MI. Following an acute myocardial infarct, necrotic cells release soluble messenger molecules, such as Damage-associated molecular patterns (DAMPs) and pro-inflammatory cytokines and chemokines, leading to the attachment and invasion of the infarct zone by immune cells - specifically neutrophils and monocytes. While the general immune cell-types and signals implicated in the sterile inflammatory response to myocardial infarction have been extensively studied, the spatiotemporal infiltration routes into the infarct zone are less well understood.

Here we aim to define the cellular immune landscape and their infiltration routes in a minimally invasive model of myocardial infarction during the acute phase using single cell resolved spatial omics technologies. To achieve this, we induce myocardial infarction by occluding the left descending coronary artery in mice. It enabled a highly reproducible and multiplexed spatial analysis of transcript expression at single-cell resolution using Molecular Cartography™ from Resolve Biosciences and highly multiplexed imaging at different time points. Afterwards, we integrate the different spatial omics measurements to create a comprehensive map of immune cell infiltration in time and space during the acute phase of the healing heart.

Preliminary data suggests that the complex interplay between the diverse cellular composition of the left ventricle (cardiomyocytes, fibroblasts, endothelial cells, vascular smooth muscle cells, pericytes) and invading and resident immune cells may hold important clues for improved treatment to prevent heart failure.

Other Biology
Omics

POSTER 409

Epstein-Barr virus genome analysis by long-read sequencing and its potential application in exploring the pathogenesis of EBV-associated diseases

Hui Chen[1], **Sia Yee Yen**[1], **Miao Xu**[2], **Robert Vaser**[3], **Mile Sikic**[3,4], **Melvin Lee Kiang Chua**[5,6], **Thomas Loh**[7,8], **Joshua Tay**[7], **Patrick Tan**[9,10,11,12,13], **Jianjun Liu**[1,14*]

1 Human Genomics, Genome Institute of Singapore, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore
2 State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Sun Yat-sen University Cancer Center, Guangzhou, China
3 Laboratory for Bioinformatics and Computational Biology, University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia
4 Laboratory of AI in Genomics, Genome Institute of Singapore, A*STAR, Singapore, Singapore
5 Division of Radiation Oncology, National Cancer Centre, Singapore, Singapore
6 Duke-NUS Graduate Medical School, Singapore, Singapore, Singapore
7 Division of Surgical Oncology (Head and Neck Surgery), National University Cancer Institute, Singapore
8 Department of Otolaryngology, Yong Loo Lin School of Medicine, National University of Singapore (NUS), Kent Ridge, Singapore
9 Genome Institute of Singapore, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore
10 Cancer and Stem Cell Biology Program, Duke-NUS Medical School, Singapore, Singapore
11 Biomedical Research Council, Agency for Science, Technology and Research, Singapore, Singapore
12 Cancer Science Institute of Singapore, National University of Singapore, Singapore, Singapore
13 SingHealth/Duke-NUS Institute of Precision Medicine, National Heart Centre Singapore, Singapore, Singapore
14 Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

Epstein-Barr virus (EBV) infects more than 90% of the world population, leading to the development of a wide range of cancers, including lymphomas, nasopharyngeal carcinoma (NPC) and gastric carcinoma. EBV strains and clinical isolates from diverse populations have been intensively analyzed using next-generation sequencing (NGS) technologies, which has greatly advanced our understanding about the evolution of EBV genome and the role of EBV genome variation in disease development, such as our recent discovery of high-risk EBV strains for NPC through a large-scale NGS analysis of EBV isolates from southern China (Xu et al Nature Genetics 2019). However, over 30% of EBV genome consists of repetitive sequences and cannot be analysed by NGS, and as a result, their contribution to EBV genome evolution and disease development are poorly understood.

In the current study, we sequenced 7 EBV commonly analyzed strains by using the long-read sequencing technologies from PacBio (sequel IIe) and Oxford Nanopore (PromethION Gyppy 5) and performed de novo assembling analysis using multiple assemblers (Hifiasm, Raven and Flye). For each of the 7 EBV strains, we were able to generate a continuous contig of sequences that represents the circular genome of EBV. We compared the newly assembled B95-8 genome to the current EBV reference genome (a B95-8/Raji hybrid) and identified several novel mutations, including 7 single-nucleotide polymorphisms (SNPs), 5 insertions and 8 deletions. In particular, the newly assembled B95-8 genome has one extra copy of internal repeat 1 (IR1) unit (3,072 bp) and the deletion of 1 terminal

(continue to page 106)

(continued from page 105)

repeat (TR) unit (512 bp). Further comparison between the newly assembled and previously published genomes of the other 6 strains also identified a good number of novel mutations that were not observed by previously NGS analyses. We also observed the considerable genetic heterogeneity among the genome assemblies of B95-8 from different human leukemia cell lines (LCLs) as well as the sequencing reads from a single LCL, particularly over the IR1 region. Lastly, we built the graphic presentation of the EBV pan-genome reference from 7 linear genome assemblies and explored its application for EBV genome analysis.

As the most comprehensive long-read sequencing analysis of EBV genomes, this study has not only established a de novo assembling-based pipeline for EBV genome analysis, but also generated a new EBV pan-genome reference that will empower the evolutionary studies of EBV genomes and the association analysis of EBV genome variations in disease development.

Spatial whole transcriptome profiling of spiny projection neurons in the mouse striatum

Alison VanSchoiack[1], **Kaitlyn Roman**[2], **David Kroeppler**[1], **Yan Liang**[1], **Kit Fuhrman**[1], **Nicholas Confuorto**[1], **Joseph Beechem**[1], **Ellen J. Hess**[2]

1 NanoString® Technologies, Seattle, WA
2 Department of Pharmacology and Chemical Biology, and Department of Neurology, Emory University School of Medicine, Atlanta, GA

Defining the molecular architecture of gene expression within the brain is critical for understanding normal brain function and for developing and testing new hypotheses toward understanding and treating neurological disease. Novel spatial transcriptomic methods enable high resolution in situ profiling of individual cell-types while maintaining the organizational context of each population profiled. Here, we use NanoString's GeoMx™ Digital Spatial Profiler in combination with the Mouse Whole Transcriptome Assay (WTA) to segment and profile subpopulations of spiny projection neurons (SPNs) in the mouse striatum. SPN subpopulations are generally divided into those that express D1 dopamine receptors and those that express D2 dopamine receptors. Recently, it was shown that these subpopulations are more complex than previously thought whereby gene expression in SPNs is, in part, mediated by their position within regional striatal subdivisions.

To analyze SPNs, 10um fresh frozen coronal sections from six C57/BL6 (WT) mice (n=3 female, n=3 male) were placed across two slides. D1 and D2 SPNs were identified using Drd1 and Adora2a RNAScope probes respectively in combination with Syto13 for visualization of all nuclei. Regions of interest (ROIs) were selected in dorsomedial, dorsolateral, and ventrolateral striatum to identify cell-type specific gene expression signatures in regions previously associated with unique functional roles within the striatum. Within each ROI, Adora2a and Drd1 positive neurons were profiled and analyzed separately with average gene detection in each population ranging from 6000-9000+ genes. In Drd1-expressing SPNs, 171 genes were differentially expressed in dorsomedial versus dorsolateral striatum, and 356 genes were differentially expressed in dorsolateral versus ventrolateral striatum (Benjamini-Hochberg padj<0.1; |FC|≥1.25). In Adora2a-expressing SPNs, 76 genes were differentially expressed in dorsomedial versus dorsolateral striatum, and 335 genes were differentially expressed in dorsolateral versus ventrolateral striatum (Benjamini-Hochberg padj<0.1; |FC|≥1.25). GeoMx data was benchmarked against bulk and single-cell RNAseq from matched samples with high concordance across all metrics, and spatial analysis yielded novel insights into striatal organization of SPNs in WT mice.

In conclusion, we have established a workflow enabling whole transcriptome data from spatially mapped SPN populations within healthy mice and demonstrate high precision and accuracy of the approach. We demonstrate functional gene expression differences between SPNs based on spatial localization within the striatum highlighting spatial context's critical importance in transcriptomic analysis. These data and workflow can be used as standards to inform future profiling studies in mouse models of neurological disorders and disease.

Other Biology
Omics

POSTER 411

Simple and fast DNA extraction for multiple sequencing technologies using Digital Fluidics

Simran Kaushal[1], **Devin Junkins**[1], **Febriana Pangestu**[1], **Rithika Raj Kumar Pradeep**[1], **Gabriella Davis**[1], **Raphael Merand**[1], **Abdul Mohammed**[1], **Will Langford**[1], **Udayan Umapathi**[1]

1 Volta Labs, Inc., Cambridge, MA

Integrating data from different sequencing platforms is the next frontier in biological advancement. Additionally, recent improvements in long-read sequencing technologies have made consistent extraction of high-molecular-weight (HMW) DNA a bottleneck. What biologists need are fast and simple DNA extraction methods. Here, we present an innovative approach for simultaneous HMW DNA extraction for multiple sequencing technologies using the Digital Fluidics system.

The Digital Fluidics system is able to extract high-quality, contamination-free HMW DNA. The system provides consistent purity and yield outputs, allowing pooling of multiple HMW extractions if necessary. DNA extracted using this approach was successfully sequenced using both long-read and short-read sequencing technologies. In a single run this approach can isolate HMW DNA from cell lines that is more than sufficient for de novo assembly and variant detection using PacBio, and early indications show that HMW DNA from Digital Fluidics contains more intact DNA molecules than manual extractions. Similarly, a single run generates sufficient HMW DNA >150 kb to reach N50s of 20-25 kb and data yields of ~15 Gb on the Oxford Nanopore Technologies (ONT) MinION sequencing platform. DNA extracted from cell lines and whole blood sequenced on one lane of the Illumina NovaSeq S4 gave over 30x genome coverage and ~90% of bases with Q ⩾ 30.

We demonstrate the production of high-quality HMW DNA compatible with sequencing on all major platforms using the Digital Fluidics system. Our system is easy to use and provides consistent results, with development efforts focusing on increasing yield and HMW DNA lengths. We believe Digital Fluidics could be crucial in advancing biological discoveries from co-sequencing of HMW DNA on multiple sequencing platforms.

Other Biology
Omics

POSTER 412

Spatial transcription profiling (Molecular CartographyTM) of the murine brain following chronic glucocorticoid exposure

Daniel Torres[1], **Matthew W. Pitts**[1], **Lucia A. Seale**[1], **Katlyn J. An**[1], **Ashley N. Hanato**[1], **Jessica L. Nicholson**[1], **Marla J. Berry**[1]

1 University of Hawaii at Mānoa, Mānoa, Hawaii

The role of chronic stress in human disease pathology has become a topic of major interest spanning multiple fields. While the stress response serves an important survival role, high levels of stress can have deleterious effects on an organism, including neurological impairments and disrupted energy homeostasis. The micronutrient selenium is essential for proper brain function and is used to synthesize a family of proteins participating in redox reactions called selenoproteins. The ability of selenium to counteract the neurological detriments caused by stress has recently begun to come to light. Conversely, preliminary data suggests that long-term administration the glucocorticoid stress hormone corticosterone negatively regulates the expression of selenoproteins. We hypothesize that this perturbation of selenoprotein expression contributes to the neurological impairments caused by stress. The aim of this study was to assess the transcriptomic effects of chronic glucocorticoid exposure in the brain of the mouse, with a focus on the selenoprotein family and selenium metabolism genes. C57 wild-type mice were administered either corticosterone or vehicle (ethanol) for four weeks via drinking water. The Molecular Cartography™ platform was used to analyze regional changes in gene expression throughout the hypothalamus and hippocampus. These data revealed the cell-type specific interactions between selenium and glucocorticoids in the brain.

Other Biology
Omics

POSTER 413

Spatial Organ Atlas: a compendium of whole transcriptomes for functional subunits of organs

Michael Rhodes, Patrick Danaher, Jason Reeves, Tyler Hether, Megan Grout, Alison VanSchoiack, Jingjing Gong, Liang Zhang, Wei Yang, Charlie Glaser, Kathy Ton, David Kroeppler, Stefan Phelan, Kit Furhman, Nicholas Confuorto, Swati Ranade, an Liang, Sarah Warren

NanoString® Technologies, Seattle, WA

The Spatial Organ Atlas (SOA) is the first repository of in-situ whole transcriptome data generated from FFPE, non-diseased organs. The spatial whole transcriptomes were generated by Digital Spatial Profiling, allowing transcriptomes that match the exact shape of the functional histological organ features.

In this first release of the SOA, data has been collected across six organs: human kidney, brain, lymph node, pancreas, colon, and mouse brain. Regions of organs with complex geometries can be profiled with the whole transcriptome in a manner that exactly matches defined histological structures. This can be accomplished even when a structure is composed of a complete interweaving of varied cell types (e.g., the alpha and beta cells of the Islets-of-Langerhans), allowing capture of pure in-situ transcriptomes of the overlapping cell types. Over 1200 whole transcriptomes were measured in the SOA. Biological interpretation is facilitated using the interactive data viewer Minerva and supplied analysis vignettes. To further enable exploration of the data, the raw and processed data (high-resolution images in OME-Tiff format, the FAST-q files, and transcript counts) are available for download. Seurat data objects and GeoMx tools data objects are also available to support analyses using Bioconductor R based tools.

In the next phase of the initiative, we are creating application examples for how organ atlasing can be used for discovering unique insights in disease biology. Additionally, we will continue to enrich the repository with new datasets for different organs as this project progresses. These whole transcriptome atlases will serve as reference for advancing our understanding of organs in development, health and disease.

For research use only, not for use in diagnostics

Other Biology
Omics

POSTER 414

Decoding the development of the human spinal cord by multi-omi

Xiaofei Li[1], Zaneta Andrusivova[2], Paulo Czarnewski[2,3], Alma Andersson[2], Christoffer Langseth[4], Daniel Gyllborg[4], Emelie Braun[5], Ludvig Larsson[2], Lijuan Hu[5], Zhanna Alekseenko[1,6], Helena Kopp Kallner[7], Elisabet Åkesson[1,8], Igor Adameyko[9,10], Mats Nilsson[4], Sten Linnarsson[5], Joakim Lundeberg[2], Erik Sundström[1]

1 Division of Neurogeriatrics, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden
2 Science for Life Laboratory, Department of Gene Technology, KTH Royal Institute of Technology, Stockholm, Sweden
3 Science for Life Laboratory, Department of Biochemistry and Biophysics, National Bioinformatics Infrastructure Sweden, Stockholm University, Stockholm, Sweden
4 Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden
5 Division of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden
6 Department of Cell and Molecular Biology, Karolinska Institutet, Stockholm, Sweden
7 Department of Clinical Sciences, Danderyd Hospital, Danderyd, Sweden
8 R&D Unit, Stockholms Sjukhem, Stockholm, Sweden
9 Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden
10 Department of Neuroimmunology, Center for Brain Research, Medical University of Vienna, Vienna, Austria

Development of the human spinal cord remains largely unknown. Here in our work, we integrate single-cell RNA sequencing, spatial transcriptomics and in situ sequencing to characterize cell fate determination and differentiation through time and across multiple regions. We identify that most major cell populations in human adult spinal cord have been specified within the first trimester. We discover the genes associated to cell fate choices and their locations. Interestingly, we find that neurogenesis and gliogenesis are not regulated in temporally separate phases in human compared to rodents, and pinpoint the similar origin shared by three glial populations. Thus, our study provides resource for human developing spinal cord and pinpoint the potential cell and gene therapeutic targets for spinal cord diseases.

Enhancer-promoted interactome drives neuronal cell differentiation and contextualizes Alzheimer's-associated polymorphisms

Other Biology
Omics

POSTER 415

Cory C. Padilla[1], **Mital Bhakta**[1], **Michal Levy-Sakin**[1], **Lisa Munding**[1], **Marco Blanchette**[1]

1 Dovetail Genomics, Scotts Valley, CA

Chromatin Immunoprecipitation, coupled with sequencing, has shaped our understanding of how transcriptional machinery interacts with genomes to facilitate gene regulation. These protein-DNA interactions are fundamental components of primary-feature epigenetic signatures that are informative across development, gene regulation, and disease research. A challenge associated with these data is that they depict protein anchors in linear space and do not consider the complex and hierarchical genomic folding associated with DNA packaged in chromatin and further compacted into a nucleus. Here we combine chromatin immunoprecipitation with Hi-C in a single assay, known as HiChIP2, to describe the interactome associated with protein anchors. By comparing the interactome of H3K27ac (enhancers), promoters (H3K4me3), and topological boundaries (CTCF) in iPSC and NSC, we can describe the different transcriptional landscape of the essential genes involved with pluripotency maintenance, NANOG and OCT4. We then investigate the enhancer interactome of NSC with regards to Alzheimer's polymorphisms identified through GWAS. We show four distinct enhancer regions associated with APOE, a highly implicated gene in Alzheimer's disease. These results posit that SNPs occurring at enhancer kilobases away from a gene promotor could impact gene regulation2. These data introduce an additional dimension to ChIP-seq experiments to more completely describe the mechanism of gene activation and provide a more robust annotation of both protein anchors to their target gene as well as disease-associated polymorphisms.

The critical role of massive testing and genomic surveillance in New England's COVID-19 response

Other Biology
Omics

POSTER 416

Sean English[1], **Michelle Cipicchio**[1], **Jonna Grimsby**[1], **Matthew DeFelice**[1], **Brendan Blumenstiel**[1], **Christine Loreth**[2], **Corey Nolet**[1], **Gina Vicente**[1], **Katie Larkin**[1], **Natasha Smith**[1], **Faye Reagan**[1], **Katelyn Messer**[1], **Erin LaRoche**[1], **Nick Fitzgerald**[1], **Zoe Mandese**[1], **Maesha Ulcena**[1], **Lisa Green**[1], **Katie Siddle**[3], **Heidi Rehm**[1], **Niall Lennon**[1], **Daniel Park**[3], **Pardis Sabeti**[3], **Bronwyn MacInnis**[3], **Stacey Gabriel**[1]

1 Genomics Platform, Broad Institute of MIT and Harvard, Cambridge, MA
2 Data Sciences Platform, Broad Institute of MIT and Harvard, Cambridge, MA
3 Viral Computational Genomics Group, Broad Institute of MIT and Harvard, Cambridge, MA

Genomic surveillance has played a crucial role in the public health response to the COVID-19 pandemic. Prior to the pandemic, pathogen sequencing activities were typically done in academic and public health labs operating at a small scale. Even earlier in this pandemic, sequencing efforts in the US were relatively small and inhibited by a lack of connection from the vast number and varied diagnostic labs in the US to sequencing, analysis and reporting capabilities needed to respond at the scale required.

Early in the pandemic, we put in place a mega-scale CLIA diagnostic testing capacity and have produced a total of over 22.4 million diagnostic tests with an average daily volume of 60,000 tests. With support from the CDC, we have developed viral sequencing capacity to enable us to sequence for genomic surveillance at scale, averaging as many as 80% of positives, depending on testing volume. Importantly, because our diagnostic lab is tightly integrated with the sequencing process, we have greatly reduced the cycle time from positive diagnostic test to sequence-based variant identification. Sequencing data is analyzed through the Terra cloud-based platform SARS-CoV-2 pipeline. The pipeline de-multiplexes each sample, aligns reads to the genome, and identifies variants for classification. Viral genomic data is deposited in NCBI/GSAID and submitted to local public health authorities.

We have generated viral sequence data for over 35,000 samples in 2021 and typically return data to CDC and relevant local public health authorities within one week of identifying a positive sample. We have used this testing and sequence capability in collaboration with the MADPH to rapidly characterize localized outbreaks, such as the emergence of the delta variant in an outbreak in Provincetown MA in July 2021, and demonstrate the pace at which Delta became the predominant variant in Massachusetts.

Finally, our tight coupling of diagnostic testing and sequencing allows routine monitoring or viral sequences for mutations that may affect test performance (PCR drop outs). This data is crucial to ensure robust test performance and is useful broadly by diagnostic labs across the country.

Spatial transcriptomics (Molecular CartographyTM) reveals regional and subclass-specific changes with age in non-neuronal cells of the mouse brain

Other Biology
Omics

POSTER 418

Cindy T.J. van Velthoven[1], **Kelly Jin**[1*], **Kimberly Smith**[1], **Jeff Goldy**[1], **Nick Dee**[1], **Bosiljka Tasic**[1], **Hongkui Zeng**[1]

1 Allen Institute for Brain Science, Seattle, WA
*Presenting author

The mammalian brain is a highly heterogeneous tissue that is composed of a diverse collection of neuronal and non-neuronal cell types that vary in function and composition across different regions of the brain. Single-cell genomics methods are rapidly expanding our definition of cell types as well as our knowledge of their properties and functions. However, how these cell types change with age is not yet fully understood. As such, comprehensive single-cell studies profiling different non-neuronal cell types in the brain will be critical to fully elucidating how the brain changes with age.

Here we present a comparison between an unbiased sampling of non-neuronal cells from young adult (p56) and aged (18mo) mice via single cell transcriptomics. Cells were isolated from 16 different regions across the whole brain. Single cell RNA-seq data was generated using 10x Genomics v3 for 444,870 aged and 223,560 young adult non-neuronal cells. After performing de novo cell clustering and label transfer from previous RNA-seq datasets, we detect a diverse range of non-neuronal subclasses, including astrocytes, oligodendrocytes, endothelial cells, microglia, smooth muscle cells, and others across all 16 regions. We detect a multitude of unique clusters within each of these subclasses based on differential gene expression, revealing a highly complex cellular structure within non-neuronal brain cells alone. We also find unique signatures within specific subclasses, such as regional diversity across astrocyte cell types, as well as a distinctive age-specific signature across multiple oligodendrocyte cell types. To investigate the age-related changes across different regions in more detail, we performed spatial transcriptomics (Molecular CartographyTM – Resolve Biosciences) experiments. Using a gene panel of 100 genes, we verified the age-related, regional specific cell type composition in 2 specific brain regions in young adult and aged mice.

As analysis of this tremendous dataset progresses, we will continue to uncover novel insights about specific non-neuronal cells and age-related gene expression changes across broad regions of the mammalian brain.

Long reads reveal epigenetic patterns from telomere to telomere

Other Biology
Omics

POSTER 419

Ariel Gershman[1], **Savannah Hoyt**[5], **Paul W. Hook**[1], **Roham Razaghi**[1], **Sergey Koren**[4], **Michael Sauria**[1], **Nick Altemose**[6], **Gina V. Caldas7**, **Glennis Logsdon**[3], **Arang Rhie**[4], **Michael C. Schatz**[1], **Rachel O'Neill**[5], **Adam M. Phillippy**[4], **Karen Miga**[2], **Winston Timp**[1]

1 Johns Hopkins University Department of Biomedical Engineering, Baltimore, MD
2 UC Santa Cruz Genomics Institute, University of California Santa Cruz, Santa Cruz, CA
3 Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA
4 Genome Informatics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD
5 Institute for Systems Genomics, Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT
6 Department of Bioengineering, University of California Berkeley, Berkeley, CA
7 University of California Berkeley, Department of Molecular and Cell Biology, Berkeley, CA

Since the initial release of the human genome two decades ago, new updates to the reference effectively covered the euchromatic fraction of the genome, but heterochromatic and complex regions were still left unfinished or erroneous. The Telomere-to-Telomere (T2T) Consortium has finished the first truly complete sequence of a human genome by completing the last 8%. This first complete human genome presents the opportunity to analyze epigenetics in previously uncharted areas. Here we present an epigenetic annotation of a complete human genome highlighting genomic regions miss-assembled in GRCh38, expanded copy number gene families involved in human disease (ie. NBPF, DUX4), and large repetitive regions including centromeric satellite DNA (CenSat).

First, we developed novel dynamic k-mer directed mapping strategies to map existing short-read datasets from ENCODE (44 ChIP-seq sets) to the newly revealed parts of the genome. The improvements offered by the T2T-CHM13 assembly improved peak calling by granting 3-21% more peak calls compared to GRCh38. But there are regions of the genome, so-called "marker deserts" that prevent even this approach from working. 6.13% of the human genome is devoid of unique k-mers and therefore is essentially unmappable using short-read data. These regions consist mostly of segmental duplications (SDs) (2.14%), CenSat DNA (2.05%) and SDs within CenSat DNA (1.3%). In these regions, long-read epigenetics approaches are necessary. By using long-read nanopore sequencing to profile DNA methylation and chromatin, we sampled 32.2M CpG sites (99.72% of total CpGs) in T2T-CHM13 compared to 28.9M CpG sites (99.3% of total CpGs) in GRCh38. Using long reads we were able to interrogate allele specific, long-range epigenetic patterns in complex satellite arrays such as the D4Z4 repeat associated with Facioscapulohumeral muscular dystrophy.

(continue to page 116)

Other Biology
Omics

POSTER 419

(continued from page 115)

Turning to the centromere, we examined the methylation pattern of the alpha satellite higher order repeat arrays in CHM13 and HG002 and revealed a distinctive region of hypomethylation, coined the Centromeric Dip Region (CDR), consistent with decreased chromatin accessibility (nanoNOMe) and enrichment of CENP-A chromatin. Additionally, we analyzed the newly assembled X chromosomes of 4 additional diverse male samples. Not only did we identify CDRs in these samples, but we noted shifts in the positioning of the CDR within the CenSat, suggesting while the presence is universal, the exact location may vary between genotypes. Our work demonstrates the feasibility of probing an entire human methylome and reveals the complex epigenetic structure within the genome's most elusive regions.

Massively multiplex single-cell 3D genome imaging with OligoFISSEQ

Other Biology
Omics

POSTER 420

Huy Q. Nguyen[1,2], **Shyamtanu Chattoraj**[1,2]

1 Department of Genetics, Harvard Medical School, Boston, MA
2 Acuity Spatial Genomics, Newton, MA

Recent advances in genome imaging and next-generation sequencing are revealing the incredible complexity and biological importance of the genome's 3D configuration. Still, many questions remain: How does 3D genome-organization relate to normal health and disease? How are chromosomes organized in aneuploidies? How do sub-structures in chromosomes affect regulation of transcription? How much do non-coding areas of the genome affect health and disease? Being able to answer these questions requires a complete spatial map of the genome. Unfortunately, directly imaging multiple loci (>5) to visualize genome 3D organization within large numbers of single nuclei has been challenging.

Towards the goal of directly visualizing entire genomes at high resolution, we developed OligoFISSEQ1. OligoFISSEQ combines Oligopaint technology for "painting" loci with complex libraries of inexpensively synthesized oligonucleotides and the multiplexing power of fluorescent barcodes to simultaneously target, visualize, and localize any number of genomic loci. In particular, the OligoFISSEQ suite of methods involve hybridizing barcoded in situ hybridization probes (Oligopaints) to chromatin targets, followed by in situ barcoded detection and decoding. The scalability of barcode sequencing means fewer than 8-cycles are required to robustly detect every human gene.

Here, we will describe OligoFISSEQ and its application for exploring genome-organization. We will show significant improvements that we have made to the published work to increase scalability in detecting higher number of targets across the genome in single-cells making it a powerful tool to trace targeted genomic structures. We recapitulate known characteristics of genome-organization, such as chromosome territories, thus, validating our technology. Additionally, we use OligoFISSEQ to trace individual chromosomes at high genomic resolution. Hierarchical clustering of these individual chromosome traces allowed identification of subgroups of cells exhibiting specific chromosome structures. To demonstrate OligoFISSEQ sensitivity, we apply OligoFISSEQ to a variety of diseased and non-diseased cell and tissue types to identify differences in genome-organization. Being able to obtain population averaged data while maintaining single-cell resolution and distinction of individual cells is essential for identifying problematic cells within a tissue and will further elucidate the links between genome-organization and disease.

1 Nguyen HQ, Chattoraj S, Castillo D, Nguyen SC, Nir G, Lioutas A, Hershberg EA, Martins NMC, Reginato PL, Hannan M, Beliveau BJ, Church GM, Daugharthy ER, Marti-Renom MA, Wu CT. Nat Methods. 2020

Other Biology
Omics

POSTER 421

The use of high-quality genome assemblies for conservation genomics

Olivier Fedrigo[1], **Jennifer Balacco**[1], **Nivesh Jain**[1], **Brian O'Toole**[1], **Tatiana Tilley**[1], **Kathleen Horan**[1], **Sadye Paez**[2], **Giulio Formenti**[1,2], **Erich D. Jarvis**[2,3]

1 Vertebrate Genome Laboratory, The Rockefeller University, New York, NY
2 Laboratory of Neurogenetics of Language, The Rockefeller University, New York, NY
3 Howard Hughes Medical Institute, Chevy Chase, MD

The Vertebrate Genomes Project (VGP) is an international and multidisciplinary project of the Genome10K consortium (G10K). The mission of the VGP is to sequence high-quality assemblies of all ~70,000 extant vertebrate species. The primary outcome of the VGP is to create a repository of high-quality genomes assemblies to not only preserve genetic information but also to provide valuable tools for conservation efforts. Because sequencing technologies evolve rapidly, VGP members are continuously testing and evaluating new technologies and protocols. We currently use a combination of PacBio Hifi, Bionano optical maps, and Arima Hi-C to obtain the most contiguous, complete, and accurate assemblies. These past 3 years our group has generated more than 150 high-quality vertebrate genome assemblies and many of them are from endangered species. Here we report results for ~150 genome assemblies sampled across the entire vertebrate tree. We will particularly focus on genomes of endangered species and will also reflect on the use of high-quality assemblies for species conservation, describing our current and future plans to contribute to conservation genomics.

Other Biology
Omics

POSTER 422

Draft genome assembly of Saguinus oedipus

Esko Kautto[1,2], **Charles Thomas Gregory**[3], **Christoph Weigel**[3], **Jadwiga Labanowska**[4], **Yue-Zhong Wu**[3], **TJ Doong**[3], **Arletta Lozanski**[3], **Ann-Kathrin Eisfeld**[3,4], **Robert Baiocchi**[3,4], **Christopher Oakes**[3,4,5], **James S. Blachly**[3,4,5]

1 Nationwide Children's Hospital, The Steve and Cindy Rasmussen Institute for Genomic Medicine, Columbus, OH
2 The Ohio State University Department of Pediatrics, Columbus, OH
3 The Ohio State University Department of Internal Medicine, Division of Hematology, Columbus, OH
4 The Ohio State University Comprehensive Cancer Center, Columbus, OH
5 The Ohio State University Department of Biomedical Informatics, Columbus, OH

Saguinus oedipus (Cotton-top tamarin; CTT) is a critically endangered new-world monkey with a habitat limited to a small area in northern Colombia. Factors leading to endangered status include rainforest destruction, illegal pet trade, and tragically, biomedical research. Because S. oedipus develops colitis and colon adenocarcinoma, tens of thousands of animals were exported as ulcerative colitis and spontaneous colon cancer models in the mid-twentieth century. Additionally, research into herpesvirus-induced cancers in the 1970s revealed that CTT inoculation with Epstein-Barr Virus (EBV) consistently yielded lymphoma that recapitulated human Burkitt lymphoma. Although direct use of CTT as a lymphoma model ended due to conservation concerns, an immortalized cell line, B95-8, was created from a CTT individual that developed aggressive lymphoma. B95-8 is still widely used in EBV research, but is difficult to obtain in many jurisdictions including the US, where CTT is listed on the Endangered Species List.

To date, no CTT reference genome exists; the only reference genome for any new world monkey is the common marmoset, Callithrix jacchus. Because a reference genome for any endangered animal has inherent value, the paucity of data for new-world monkeys generally, and because a reference genome for a historically medically-important species could shed new light on past discoveries without incurring additional risk to the animal, we set out to assemble the first draft genome for S. oedipus using B95-8 material as a primary source.

We first karyotyped B95-8 and found that its ideogram was a good match to S. oedipus with minimal cytogenetic alterations. The final karyotype was 46,XY,+Y,ins(3),-14,der(16)t(14;16)[20], with 14 joined end-to-end with one copy of 16, which can be trivially resolved in assembly.

We then pursued a nearest-relative reference-guided approach, using a combination of Illumina, Oxford Nanopore (ONT), and BioNano optical mapping data. Production of Hi-C and PacBio HiFi data are in preparation. We generated an estimated 40X coverage with 2x150 reads and assembled contigs with ABySS. We mapped these to the marmoset genome in order to assign each a chromosome number for later reassembly. We next polished ONT data with Illumina reads to generate longer scaffolds. Finally, we generated 2825 Gbase (est. 942X) of high-quality optical maps, with N50 of ~255 kbase (before de novo assembly). Final assembly and annotation of the draft genome will be completed after gathering additional Hi-C and long-read data.

Application of spatial transcriptomics (Molecular CartographyTM) to investigate disruption of hepatic zonation in 2,3,7,8-tetrachlorodibenzo-p-dioxin elicited progression of steatosis to steatohepatitis with fibrosis

Other Biology
Omics

POSTER 423

Rance Nault[1]

1 Michigan State University, East Lansing, MI

Cellular heterogeneity and structural organization are essential for normal liver homeostasis and its response to drugs and chemicals. Chemicals such as the persistent environmental contaminant and potent aryl hydrocarbon receptor (AHR) agonist, 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD), elicit distinct dose-dependent spatially resolved patterns of tissue damage. While histological assessment initially identified dose-dependent portal zone effects that expanded towards the central zone following oral gavage with TCDD, bulk and single-nuclei transcriptomic studies suggested initial responses occurred in the central zone leading to a loss of zonal identity that contributed to widespread tissue damage. To further investigate the dose-dependent spatially resolved transcriptomic effects of TCDD, the Resolve Biosciences' Molecular CartographyTM platform was used to characterize the expression of 97 hepatic genes in male mice orally gavaged with sesame oil vehicle, 0.3, 3, or 30 µg/kg TCDD every 4 days for 28 days. Markers of portal (Cyp2f2) and central (Glul) hepatocytes showed expected zonation patterns at 0 – 3 µg/kg, that was markedly disrupted at 30 µg/kg. Central localized genes involved in carbohydrate metabolism (Gulo), antioxidant defense (Gstm3, Gsta), and lipid uptake (Cd36, Vldlr) co-localized with Cyp2f2 at the highest dose. Examination of the dose-dependent central zone expression pattern for Cyp1a1, Fmo3, Nqo1, and Xdh is consistent with higher level Ahr expression in the central zone resulting in their induction initially occurring in the central zone and progressing to the zone. Furthermore, the emergence of a highly expressing Gpnmb and Trem2 sub-population of macrophage that also expressed Adgre1 and Cd5l markers is consistent with a reported non-alcoholic steatohepatitis (NASH) associated macrophage (NAMs) population. At 30 µg/kg TCDD, Tagln expressing hepatic stellate cells (HSCs) co-localized with activated HSC marker Acta2, and the expression of collagens Col1a1, Col1a2, and Col3a1 within periportal fibrotic tracts. Collectively, spatial transcriptomics indicated a global disruption of zonal identity within hepatocytes driven by increased expression of genes characteristic of the central zone in the portal regions. These results suggest the loss of zonal identity may be a contributing factor in the hepatotoxicity and progression of steatosis to steatohepatitis with fibrosis elicited by TCDD. This study was supported by be the National Institutes of Health Science (NIEHS R01 ES029541), NIEHS Superfund Research Program (NIEHS SRP P42 ES004911), and National Human Genome Research Institute (NHGRI R21 HG010789).

Illuminating mammalian brain taxonomy with new tools and strategies from Veranome Biosystems' spatial biology toolkit

Other Biology
Omics

POSTER 424

Brian S. Hilbush[1], **Yun-Ching Chang**[1], **Erica Teo**[2], **Edita Aliwarga**[2], **Li Zhen**[2], **Debjit Ray**[1], **Bongjun Son**[1], **Chloe Kim**[1], **Marc Glazer**[1], **Alexander Smith**[7], **Paul Kenny**[7], **Kristin G. Beaumont**[3,4], **Robert Sebra**[3,4,5,6]

1 Veranome Biosystems, Mountain View, CA
2 Veranome Biosystems, Nucleos, Singapore
3 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY
4 Center for Advanced Genomics Technology, Icahn School of Medicine at Mount Sinai, New York, NY
5 Black Family Stem Cell Institute, New York, NY
6 Sema4, a Mount Sinai Venture, Stamford, CT
7 Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY

The complex sensory, motor, and physiological response properties of the mammalian brain are achieved through well-defined neuronal circuitry and the interactions of hundreds of neuronal and non-neuronal cell types found in highly distinct and spatially organized regions. In the nervous system, definitive cell type and cell state classification require a combination of morphological and molecular features, connectivity patterns, and electrophysiological and functional properties that have historically been obtained serially through use of independent methods. Here, we have generated a large corpus of data spanning cortical, thalamic, and hindbrain regions from the mouse brain using Veranome Biosystems' VSA-1 platform. The multi-omic analysis of nearly 250,000 cells provides a basis for building an integrative taxonomy that captures several biological dimensions in situ and in parallel, including transcriptomic profiles, morphological features and spatial networks. In addition to cell type, we have mapped changes in transcriptional states in forebrain regions following acute administration of drugs of abuse. Using Veranome's highly multiplexed FISH assay, transcriptional profiles were generated for hundreds of genes and spatial maps resolved at the single cell level. Spatial transcriptomic analysis and multi-omic visualization were done with data imported into Veranome's desktop software. Additional analyses were performed following export of data into Seurat v4.0 or Scanpy 1.8.1 and with custom algorithms. With these tools, we surveyed over 50 regions of interest (typically 1 to 4 mm2), cataloguing an array of cell type-specific clusters and differentially expressed genes. By employing imputation strategies, we extended the dimensionality of any single gene panel and classified cells based on annotations either from linked Veranome panels or scRNAseq reference data. Multi-omic analysis, gene co-expression network and spatial proximity analyses further refined taxonomic classification for some neuronal subtypes. We identified potentially novel thalamic interneuron and projection neuron subtypes based on transcriptomics and spatial data. Results will be presented on the gene ensembles, morphological features and spatial neighborhood metrics that define our classification approach. Our data demonstrate that the VSA-1 is a powerful spatial multi-omic platform for interrogation of multiple biological dimensions required for taxonomic classification. The results indicate that improved cell type and cell state classification accuracy can be achieved by application of imputation methods and spatial pattern detection algorithms. Application of Veranome's toolkit will advance the building of an integrative taxonomy of the mammalian brain and enable researchers to track cellular responses to drugs, pathogens and disease progression.

Mapping transcriptome heterogeneity in the adult central nervous system with single-cell resolution using Molecular Cartography™

Other Biology
Omics

POSTER 425

Matthew Holt[1,2], **Araks Martirosyan**[2], **Francisco Pestana**[2], **T. Grant Belgard**[3], **Pascal Bielefeld**[4], **Carlos Fitzsimons**[4], **Dietmar Schmucker**[1,5]

1 Instituto de Investigação e Inovação em Saúde (i3S), University of Porto, Porto, Portugal
2 VIB-KU Leuven Center for Brain and Disease Research, Leuven, Belgium
3 The Bioinformatics CRO, Orlando, FL
4 Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, Netherlands
5 Life & Medical Sciences Institute, University of Bonn, Bonn, Germany

The central nervous system (brain) is arguably the most complicated organ in mammals, consisting of a multitude of major cell types (including neurons, astrocytes, oligodendrocytes and microglia), whose close functional interactions are responsible for generating thoughts and ideas, and ultimately controlling behaviors.

The recent explosion in single cell RNA-sequencing (scRNA-seq) has revolutionized our understanding of the molecular diversity among cell types in the brain, providing crucial insights into the variety of cell-type specific functions needed to produce a working nervous system. Unfortunately, however, critical information on the spatial relationships between these cell types, both during development as well as in response to injury and/or disease, is usually lost because of the need to dissociate tissue prior to sequencing. The ability to map gene expression data from scRNA-seq back to intact tissue is, therefore, essential to further our understanding of brain function.

Here, we describe our work, using a highly multiplexed in situ hybridization method (Molecular CartographyTM) developed by Resolve Biosciences, to produce a high-resolution map of the major cell (sub)types in the adult mouse cortex and hippocampus, which reveals novel aspects of brain architecture. Using the example of traumatic brain injury (TBI), a common cause of death and long-term disability in adults, we further demonstrate the power of this technique by systemically assessing how a severe cortical impact disrupts the hippocampus and surrounding tissues, revealing a concerted pattern of gene expression changes in individual cell types, as well as changes in spatial positioning, which may be exploitable in therapeutic settings.

Enhanced nano-bio interaction enables deep plasma proteomics at scale with increased precision and depth of coverage

Other Biology
Omics

POSTER 426

Shadi Ferdosi[1], **Tianyu Wang**[1], **Moaraj Hasan**[1], **Jessica Chu**[1], **Eltaher M. Elgierari**[1], **Xiaoyan Zhao**[1], **Martin Goldberg**[1], **Asim Siddiqui**[1], **Serafim Batzoglou**[1], **Omid C. Farokhzad**[1], **Daniel Hornburg**[1]

1 Seer, Inc., Redwood City, CA

To overcome limitation of deep plasma proteomics in large cohorts, we have developed a fast and scalable technology that employs intricate protein-nano interactions. Introducing a nanoparticle (NP) into a biofluid such as blood plasma leads to the formation of a selective, specific, and reproducible protein corona at the nano-bio interface driven by the relationship between protein-NP affinity, protein abundance and protein-protein interactions. We previously demonstrated that this process, incorporated within the Seer Proteograph™ Product Suite, offers superior performance in terms of depth, breadth, precision, and throughput compared to conventional deep workflows. The ratio of plasma-to-nanoparticles determines the competition between proteins for binding surface, which plays an important role in protein corona composition and can be optimized to enhance and differentiate protein selectivity. Here we investigate effects of different conditions on protein corona composition enabling enhanced performance of Proteograph.

We have investigated compositional changes of protein coronas from 5 NPs with blood plasma at different ratios. Samples were analyzed with timsTOF Pro mass spectrometry and UltiMate3000 Dionex LC system using 30min DIA runs. We evaluated depth, dynamic range, coverage, and precision of quantification at a wide range of concentrations for each NP.

By limiting the available binding surface of NPs and increasing the binding competition, we are able to identify 20 – 60% more proteins on the surface of each NP. Moreover, by increasing the competition the proteins are more reproducibly identified and quantified across the replicates of the same NP. In addition, protein selectivity was enhanced, leading to improved coverage of plasma proteome when using multiple physiochemically distinct NPs. In summary, NP panels with optimized workflow, capture a large and diverse set of proteins and biological pathways based on their specific physicochemical makeup.

Exploring microbial biodiversity through whole genome analysis using ResolveDNA™ Microbiome

Other Biology
Omics

POSTER 427

Durga M. Arvapall, Isai Salas-González, Jon S. Zawistowski, Jeff G. Blackinton, Mark Hurt, Gary L. Harton, Jason A.A. West

BioSkryb Genomics, Inc., Durham, NC

Genomic surveillance of microbes is crucial to understand the genetic contributions to host/bacteria interactions. In contrast to metagenomic approaches, whole genome sequencing (WGS) of single bacteria is required to fully reveal the diversity and evolutionary trajectory of the microbiome.

Here, we utilized the ResolveDNA™ Microbiome kit to generate high quality genome assemblies from two bacterial species. The workflow utilizes Primary Template-directed Amplification (PTA), a novel technology for low-input and single cell genome amplification with unprecedented coverage uniformity. Gram-negative (E. coli), gram-positive (B. subtilis) single bacterial cells and mixed bacteria (E. coli & B. subtilis) were sorted using fluorescence-activated cell sorting (FACS). The single sorted bacteria were lysed and amplified using the ResolveDNA™ Microbiome kit with amplification yields between 50-200 ng and average amplicon size of 1000 bp. ResolveDNA™ sequencing libraries were constructed, followed by Illumina MiniSeq sequencing. The reconstructed genome assemblies approximated the empirically estimated genome sizes for each species. Additionally, by employing an unbiased de novo phylogenetic approach we quantitatively validated the high-quality of the assemblies and calculated that the assembly completeness was over 95% for each single bacterium sequenced. Collectively, the data provides evidence that minute numbers of bacteria can be separated, amplified and prepared for next generation analysis (NGS) analysis.

Having successfully analyzed single bacterial genomes, currently we are expanding these findings to computationally deconvolve multi-species bacterial cocktails and to generate datasets from environmental and clinical microbiome samples. We have demonstrated here that ResolveDNA™ Microbiome platform is robust to reconstruct high quality genomes for gram-negative and gram-positive bacteria at single-cell resolution, which allows for analysis of complex microbial populations, including those not culturable.

Correlated expression of protein and RNA using bulk and single-cell proteogenomics reveals a unique molecular signature in human polarized Th1 cells

Other Biology
Omics

POSTER 428

Dipesh Risal, Tse Shun Huang, John Ma, Jeanne Elia, Kevin Taylor, Miguel Tam

BioLegend, San Diego, CA

Bulk RNA sequencing (RNA-seq) has become a widely used tool to measure gene expression whereas flow cytometry is considered a gold standard to study protein expression at a cellular level. Recent developments in single-cell sequencing technologies have combined the analysis of RNA and protein in a technique called cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq). This method utilizes antibodies conjugated to oligonucleotides, available as TotalSeq antibodies. These oligo-conjugated antibodies have now been incorporated in a bulk RNA-seq workflow to simultaneously obtain RNA and protein data, described as bulk epitope and nucleic acid sequencing (BEN-seq). The workflow renders a complete proteogenomic profile of the sample, massively increasing the number of protein targets that can be simultaneously studied compared to flow cytometry. To demonstrate the utility of this approach, we performed CITE-seq and BEN-seq with 280 TotalSeq antibodies on Th1 polarized human peripheral blood mononuclear cells.

Although T helper 1 cells have been studied for many years, the availability of novel tools and workflows such as BEN-seq allows for an unprecedented characterization of this important cell population. Thus, we found a unique molecular profile, confirmed by both methods, demonstrating how protein analysis can be used to enrich bulk RNA-seq datasets. We also expanded on the BEN-seq workflow to optimize the use of TotalSeq antibodies and demonstrated the robustness of bulk protein detection by sequencing, illustrating how biological and technical variables can be controlled for optimal protein expression profiling.

Spatial whole transcriptome profiling of mouse organogenesis

Other Biology
Omics

POSTER 429

Stephanie Zimmerman[1], **Alexa Lasley**[1], **Michael Rhodes**[1], **Kit Fuhrman**[1], **Margaret Hoang**[1], **Nicholas Confuorto**[1], **Joseph Beechem**[1]

1 NanoString® Technologies, Inc., Seattle, WA

Mammalian development is driven by key transcriptional regulators that are exquisitely specified in time and space. Insight into the transcriptional programs driving organogenesis have traditionally come from low-plex in situ methods and bulk or single cell transcriptomics of dissociated embryos. However, recent technological advances have made it possible to profile whole transcriptomes of defined regions of interest while retaining their spatial context. We used NanoString's GeoMx® Digital Spatial Profiler and the Mouse Whole Transcriptome Atlas to comprehensively profile anatomical substructures of nine organs from whole mounted fixed frozen mouse embryos (E9, E11, E13, and E15). Whole transcriptomes were collected from regions corresponding to substructures of the developing heart, lung, trachea, esophagus, stomach, duodenum, midgut, pancreas, and kidney or their precursors. Similar cell types across organs (e.g. epithelium or mesenchyme) showed both shared and organ-specific gene expression profiles, and we identified both known and novel cell-type specific marker genes and traced the dynamics of their expression. Transcription factors were among the most differentially expressed genes in time and space. We characterized the expression pattern of key transcription factors driving gut and cardiac development and identified novel spatially variable and substructure-specific expression of these transcription factors and their target genes. These results provide a spatial and temporal atlas of the transcriptional programs governing key cell fate decisions during development.

Exploiting novel long noncoding RNAs as pharmacological targets for treating Alzheimer's disease using an integrative informatics approach

Other Biology
Omics

POSTER 430

Yuji Zhang

Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, MD

As the sixth leading cause of death in the United States, Alzheimer's disease (AD) currently produces dementia in 5.8 million U.S. citizens and this number will increase to 13.8 million by 2050. However, over 200 clinical trials documented in the Clinicaltrials.gov database have not yielded therapies to prevent or slow disease progression. Accumulating evidence suggests that ADs are highly complex diseases involving dysregulation of not only protein-coding RNAs but also long non-coding RNAs (lncRNAs). Thousands of lncRNAs are expressed specifically in the brain with precisely regulated temporal and spatial expression patterns, although their specific influences in AD remain be largely unknown. Recently, we developed an integrative informatics framework to reannotate and characterize disease-associated novel lncRNAs in both melanoma and heart disease models. Furthermore, our recent studies and a few others suggested that in silico-based drug repositioning strategies can greatly expand our understanding of drug-drug target interaction and lead to novel opportunities for rational drug repositioning. In this work, we propose to investigate the potential roles of lncRNAs by examining their expression patterns across tissues in different brain regions of AD patients. Specifically, we developed an integrative informatics approach aiming to investigate the potential druggability of AD-altered lncRNAs (ADALs) by integrating three AD RNA sequencing (RNAseq) cohorts sponsored by the Accelerated Medicine Partnership (AMP) Program. First, we tailored our harmonization approach to reannotate 778 RNAseq samples across seven brain regions. Second, we employed our in-silico drug repositioning approach to identify and prioritize drug candidates targeting these ADALs though both network topology and weighted co-expression network analysis. In total, we identified and characterize 2865 tissue-specific ADALs across seven brain regions. Based on their computationally inferred perturbational profiles using the Library of Integrated Network-based Cellular Signature (LINCS) dataset, these ADALs were then prioritized using our integrative weighted and co-expression network analyses. Finally, 142 ADALs were prioritized as potential drug targets for our future independent assessment by both registered clinical trial data and AD medical expert knowledge. Our integrative informatics approach for AD drug development will greatly accelerate the underpinning of AD by repositioning known perturbagens targeting ADALs for novel AD therapy. The outcome will benefit the biomedical community by guiding future targeted experimental designs and generating testable hypotheses.

Other Biology
Omics

POSTER 431

Development of MaxPlex Capture, a high-throughput, highly multiplexed target enrichment of human exome

Steve Oh, **Elaine Maggi, Michael Bocek, Leonardo Arbiza, Rebecca Nugent, Derek Murphy, Esteban Toro, Siyuan Chen**

Twist Bioscience, South San Francisco, CA

Target enrichment through DNA hybridization-capture reduces the costs and complexity of analysis by focusing the interrogation of a DNA library to the specific genomic regions of interest. The development of targeted whole exome capture panels has allowed for sequencing output and analysis to be directed at just 1% of the genome associated with protein coding regions. This reduced burden on sequencing throughput, in combination with sample indexing strategies, has allowed for multiple samples to be pooled on a single sequencing run.

With the increase in capacity afforded by the newest generation of sequencing instruments, an increased number of samples can be combined on a single sequencing run, further reducing both costs and processing time. In 2019, to further streamline upstream sample processing, Twist Bioscience introduced modifications to our whole exome sequencing workflows that allowed for single-day, multiplexed capture of up to 16 Core Exomes.

Since these innovations, Twist has developed an improved capture system that further expands the multiplexing capability of our target enrichment system. Using a newly optimized capture panel, Exome 2.0, we demonstrate this new capability through the simultaneous capture of 96 genomic libraries. Here we present data generated on an Illumina NovaSeq 6000 sequencing platform which shows equivalent performance using the MaxPlex capture system when compared to our standard mid-plexed protocols.

These results highlight the capability of the MaxPlex capture approach and further demonstrate the potential of a high-throughput whole exome sequencing solution.

Other Biology
Omics

POSTER 433

A streamlined workflow for high-throughput, multiplexed HiFi sequencing of microbial genomes

Meredith Ashby, **Aaron Wenger, Joan Wilson, Shreyasee Chakraborty, Davy Lee, Christopher Dunn, Ivan Sovic, Zev Kronenberg, Heather Ferrao**

Pacific Biosciences (PacBio), Menlo Park, CA

HiFi sequencing on the PacBio Sequel II and Sequel IIe Systems produce accurate, long reads which – together with novel algorithms – have brought new levels of contiguity, completeness, and accuracy to the assembly of large genomes. In principle, HiFi reads are similarly beneficial for microbial genome assembly, but obstacles in library preparation workflow, library cost, and assembly algorithm (such as plasmid assembly) have limited use by microbiologists. Here, we introduce a new library prep workflow and assembly algorithm to extend the utility of HiFi reads to microbial genome assembly. The optimized library prep workflow combines steps, eliminates the need for strict size selection, and shortens reaction times to reduce workflow time to 6 hours. The assembly algorithm uses strict read-to-read overlaps enabled by HiFi read accuracy to resolve repeats. It uses a two-stage approach: a first to assemble chromosomes and a second to recover short, high-coverage plasmids.

To evaluate the approach, we generated a pool of HiFi libraries for 96 microbial samples with total genome size of 375 Mb. The protocol as evaluated with microbes relevant to pathogen surveillance including common foodborne pathogens (Listeria, Salmonella) and species often seen in hospital settings (Klebsiella, Staphylococcus). The microbes represent a range of genome sizes, assembly complexity, GC content, chromosome counts, and plasmid content. DNA samples were sheared to 7 kb - 10 kb, barcoded libraries were prepared with the optimized workflow, samples were pooled, and the pool was sequenced on one SMRT Cell 8M on the Sequel IIe System.

Reference quality de novo microbial assemblies with 5 contigs or fewer for all samples. Typical chromosome assembly quality was Q50, measured as concordance to reference assemblies. Nearly all plasmids were recovered, including those shorter than 5 kb which are often lost in workflows that require strict size selection. Taken together, the new method provides a high-throughput, low-cost approach to generate reference quality microbial genomes. similarly beneficial for microbial genome assembly, but obstacles in library preparation workflow, library cost, and assembly algorithm (such as plasmid assembly) have limited use by microbiologists. Here, we introduce a new library prep workflow and assembly algorithm to extend the utility of HiFi reads to microbial genome assembly. The optimized library prep workflow combines steps, eliminates the need for strict size selection, and shortens reaction times to reduce workflow time to 6 hours. The assembly algorithm uses strict read-to-read overlaps enabled by HiFi read accuracy to resolve repeats. It uses a two-stage approach: a first to assemble chromosomes and a second to recover short, high-coverage plasmids.

(continue to page 130)

(continued from page 129)

To evaluate the approach, we generated a pool of HiFi libraries for 96 microbial samples with total genome size of 375 Mb. The protocol as evaluated with microbes relevant to pathogen surveillance including common foodborne pathogens (Listeria, Salmonella) and species often seen in hospital settings (Klebsiella, Staphylococcus). The microbes represent a range of genome sizes, assembly complexity, GC content, chromosome counts, and plasmid content. DNA samples were sheared to 7 kb - 10 kb, barcoded libraries were prepared with the optimized workflow, samples were pooled, and the pool was sequenced on one SMRT Cell 8M on the Sequel IIe System.

Reference quality de novo microbial assemblies with 5 contigs or fewer for all samples. Typical chromosome assembly quality was Q50, measured as concordance to reference assemblies. Nearly all plasmids were recovered, including those shorter than 5 kb which are often lost in workflows that require strict size selection. Taken together, the new method provides a high-throughput, low-cost approach to generate reference quality microbial genomes.

Precision metagenomics using a hybrid assembly for classification of Shiga toxin-producing Escherichia coli in enriched agricultural water

Meghan Maguire, Julie A. Kase, Andrea Ottesen, Padmini Ramachandran, Mark K. Mammel, Dwayne Roberson, Tim Muravanda, Sandra Tallent, Eric W. Brown, Marc Allard, Steve Musser, Narjol González-Escalona

U.S. Food and Drug Administration (FDA), Silver Spring, MD

Introduction: Culture-independent metagenomic sequencing of enriched agricultural water could expedite detection and virulotyping of Shiga toxin-producing Escherichia coli (STEC). We previously determined the limit of assembly of a complete fragmented metagenome-assembled genome (MAG) for O157:H7 in enriched agricultural water using long reads (nanopore) was 105 CFU/ml.
Purpose: We aimed to establish the limits of detection and assembly for STECs in enriched agricultural water by Illumima Miseq sequencing technology alone and in hybrid assembly with nanopore long read sequencing.

Methods: STEC-negative agricultural water was enriched overnight according to the FDA BAM Chapter 4A and 1 ml aliquots were artificially contaminated with 1 ml of a 109 – 104 CFU/ml dilution of pure, overnight E. coli EDL933 culture (6 samples). Short reads were generated using the Illumina Miseq system with three samples run in duplicate per cartridge and de novo assembled using CLC Genomics (Qiagen). Long reads for the same samples generated in a previous study were used for the hybrid assemblies with Spades v3.13.1 and OPERA-MS v19.07.01.

Results: The MiSeq sequencing output (1.6-2.8 Gb per sample) generated over 11,000 contigs for each sample. For enriched agricultural water samples spiked with 107 CFU/ml and above, a complete fragmented O157:H7 MAG was achieved. The O157:H7 serotype and all virulence genes present in the spiked strain were identified. 106 CFU/ml and lower spiked concentrations produced incomplete MAGs with missing virulence genes and did not identify the O157:H7 serotype. Hybrid assembly using the short and long reads for the same samples produced variable results. None of the hybrid assemblers was able to recover a completely closed O157 MAG from any sample even with the highest inoculation level (108 CFU/ml).

Significance: Precision metagenomic sequencing can be used to generate fragmented or complete closed MAGs, but hybrid assemblies require higher coverage or sequencing depth from short read technologies.

Other Biology
Omics

POSTER 435

Spatial multi-omics resource showcasing technologies and software

Sergio Triana[1], **Shahraz Mohammad**[1], **Kristina Zuza**[2], **Liz Tuck**[3], **Bishoy Wadie**[1], **Sharath Menon**[1], **Elena Buglakova**[1], **Jovan Tanevski**[4], **Aurelien Dugourd**[4], **Giovanni Palla**[5], **Luca Marconato**[6], **Ilia Kats**[6], **Oliver Stegle**[6], **Mario Scheel-Werner**[7], **Julio Saez-Rodriguez**[4], **Fabian Theis**[5], **Omer Bayraktar**[3], **Jan Siemens**[2], **Theodore Alexandrov**[1]

1 Structural and Computational Biology Unit, European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom
2 Institute of Pharmacology, University of Heidelberg, Heidelberg, Germany
3 Wellcome Sanger Institute, Hinxton, Cambridge, United Kingdom
4 Institute for Computational Biomedicine, Medical Faculty, Bioquant, University of Heidelberg, Heidelberg, Germany
5 Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg, Bavaria, Germany
6 Genome Biology Unit, European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom
7 Resolve Biosciences GmbH, Monheim am Rhein, Germany

Recently emerged spatial omics technologies provide unprecedented insights into spatial aspects of gene expression, protein localization, and metabolite concentrations in tissues. Now, the increasingly recognized challenge lies in integrating layers of individual spatial omics into a multi-omics molecular inventory describing cellular machinery and its spatial co-organization across transcription, translation, and metabolism. Advancing in this direction requires both novel multi-omics experimental methods and novel computational methods able to link information across omics, molecularly and spatially. However, limited availability of data impedes computational developments, the bottleneck particularly prominent for the latest generations of spatial omics.

Here, we are aiming to bridge this gap with two-fold efforts. First, we are providing public data combining complementary technologies for spatial transcriptomics: highly-multiplexed RNA smFISH (Molecular CartographyTM, Resolve Biosciences) and spatial in-situ sequencing (Cartana) as well as spatial metabolomics by MALDI-imaging mass spectrometry from consecutive sections of the mouse brain. Second, we are showcasing how this data can be interrogated by the recently emerged software for omics, multi-omics, and spatial omics.

The wildtype mouse brain was used as a typical model for spatial omics. The 10 µm coronal cryosections at Bregma -1.82 captured anatomical structures containing various cell types including hippocampus and caudoputamen. First, we analyzed consecutive sections by in-situ sequencing (Cartana) and spatial metabolomics by MALDI-imaging mass spectrometry. For in-situ sequencing, we selected a panel of cell-type-specific genes as well as metabolic genes. For spatial metabolomics, we detected small molecules with the spatial resolution of 10 µm. Second, we analyzed consecutive sections by

(continue to page 133)

Other Biology
Omics

POSTER 435

(continued from page 132)

Molecular CartographyTM (Resolve Biosciences) and by spatial metabolomics. The data was quality-controlled, molecularly annotated for spatial metabolomics, and exported into a common format. Third, we showcased the recently developed software for spatial omics. We show statistical analysis performed by squidpy, a tool for spatial single-cell omics in Python. We show how METASPACE, a cloud platform for spatial metabolomics, helps data interpretation and visual overlay. For Molecular CartographyTM, we showcase the latest capacities of the software for molecular cartography. Finally, we demonstrate how molecular integrative methods MISTy and COSMOS that leverage biological knowledge can help link spatial maps of gene expression and spatial metabolomes.

The data will be publicly available under the CC-BY license. All demos will be reproducibly shared as Jupyter or Google Collaborative notebooks. The uniqueness of the provided data supplied with demos will facilitate rapid uptake and data re-use by computational biologists, thus boosting the field of spatial multi-omics.

Cortical inhibitory neurons exhibit cell-type–specific maturation programme

Other Biology
Omics

POSTER 436

Josephine Fisher[2,4], **Marieke Verhagen**[1], **Rani Moors**[1], **Monika Moissidis**[2], **Da Mi**[3], **Oscar Marin**[2], **Lynette Lim**[1]

1 Vlaams Institute of Biotechnology, University of Leuven (VIB-KUL), Center for Brain and Disease, Leuven, Belgium
2 King's College London, Center for Developmental Neurobiology, London, United Kingdom
3 Center for Life Sciences & IDG/McGovern Institute for Brain Research at Tsinghua University, Tsinghua University, Beijing, China
4 Francis Crick Institute, London, United Kingdom

Single cell RNA sequencing (scRNAseq) studies in the adult mammalian cortex have revealed neuronal diversity at unprecedented resolution. While similar approaches have been used to examine pseudotime trajectories of nerve cells as they mature, most of these studies rarely combine spatial and anatomical information to link key brain development events to the respective transcriptional programmes. This is exceptionally critical for inhibitory interneurons in the cortex since their maturation programme involves a long and protracted tangential migration, following laminar allocation, before arriving at their final cortical layer. In this study, we examine the transcriptional changes of over 9000 purified inhibitory somatostatin (Sst) positive neurons from the mouse cortex over 3 developmental stages, E16.5, P1, and P5. We identified 3 subclasses of Sst cells, long-range projecting (LRP), Martinotti interneurons (MC) and nonMartinotti interneuron (NMC) that could be further divided into 12 types. Using random forest models, we were able to reannotate Sst neurons from other published scRNAseq developmental datasets and found 11/12 types, highlighting that previous work have largely underestimated Sst neuronal diversity. Using metaneigbour analysis, developmental Sst types could be matched to adult counterparts by similar transcriptomics signatures. To examine the timing of laminar allocation of each inhibitory subtype, we performed spatial transcriptomics of 94 genes by Molecular CartographyTM (Resolve Biosciences) at P5 cortex, allowing us to link Sst cell type identity to their laminar location. We found that at this stage, the laminar allocation of some Sst subtypes differ greatly from their adult distribution, where others remain highly similar, suggesting that timing of key developmental milestones depends on cell identity. Collectively, combining scRNAseq with spatial molecular cartography, we identify both shared and unique molecular programmes critical for the postmitotic maturation of inhibitory neurons.

Using Molecular CartographyTM to reveal the defects in Alagille syndrome hepatocyte to cholangiocyte reprogramming

Other Biology
Omics

POSTER 437

Robert Schwartz[1], **Sanjay Subramanian**[2], **Kashish Chetal**[2], **Kari Huppert**[2], **Vasuretha Chandar**[1], **Saloni Sinha**[1], **Yaron Bram**[1], **Nathan Salomonis**[2], **Stacey S. Huppert**[2]

1 Weill Cornell Medicine, New York, NY
2 Cincinnati Children's Hospital, Cincinnati, OH

The liver is alone among solid organs in its ability to fully regenerate its mass and function following extensive injury. Mouse studies using genetic reporters to trace the hepatocyte or cholangiocyte (biliary epithelial cell) lineages reveal no evidence of contribution from a reserve stem cell population for hepatic cell homeostasis or liver regeneration. An alternative process to stem cell-mediated regeneration is transdifferentiation or adaptive cellular reprogramming by which a cell changes its identity. Using a mouse model of severe bile duct loss, we have demonstrated that transdifferentiation of hepatocytes to cholangiocytes in the liver can build a biliary system that failed to form during development. The hepatocyte-derived biliary system enables resolution of cholestasis and reversal of fibrosis. Unfortunately, the process of hepatocyte to cholangiocyte transdifferentiation is impaired in patients with Alagille syndrome (ALGS) wherein a characteristic feature is bile duct paucity.

Alagille syndrome (ALGS) is a multisystem developmental disorder caused by autosomal dominant mutations in JAGGED1 (95%) or NOTCH2 (1-2%), which encode key components of the Notch signaling pathway. Here, we present spatial transcriptomic data on a Jagged1 haploisufficient mouse model of ALGS with single cell resolution, obtained with the Moelcular Cartography™ platform from Resolve Biosciences. Resolve's platform was used to assess 100 transcripts in the liver of both the Jagged1 haploinsufficient mouse experimental model of ALGS and human ALGS patients Our data revealed that even though hepatocytes partially enter the transcriptional program of cholangiocytes, they are unable to fully implement it. The Molecular Cartography™ technology allowed us to assess the spatial relationship of niche and hepatocyte to cholangiocyte transdifferentiating cells and to computationally infer required receptor-ligand interactions as well as transcriptional regulators.

Developmental change of cell types in the mouse olfactory bulb cell types detected by Molecular Cartography™

Other Biology
Omics

POSTER 438

Yu Ron[1], **Ai Fang**[1], **Rahul Garg**[1], **Seth Malloy**[1], **Yongfu Wang**[1]

1 Stowers Institute for Medical Research, Kansas City, MO

The mouse brain undergoes significant changes during postnatal development thereby having a profound impact on its function. The mouse olfactory system matures within three weeks after birth. During this period, the olfactory bulb changes from a plastic state, that allows the remodeling of incoming axons from the sensory neurons to a less permissive state. It is not still unknown whether this process is associated with alterations in cell types, their transcriptomic profiles, or both. In this study, we have conducted single cell RNASeq studies to profile the olfactory bulb cell types during development. Furthermore, to further evaluate the changes in cell types and transcriptional profiles during development, we have selected a list of 100 genes, including the markers of various cell types, to perform Molecular Cartography™ (Resolve Biosciences) analysis. Using the technology, we obtain quantitative measurements of the mRNA expression levels of these the various cell types. Based on this information, we identify the specific cell types that undergo changes during the first week of postnatal development.

Mapping the crosstalk between microglia and neuronal cells in human brain organoids to understand brain development and neurological disorders

Other Biology
Omics

POSTER 439

Florent Ginhoux[1], **Tatsuya Kozaki**[1], **Oniko Liani**[1], **Sean Kim**[2], **Sarah Murphy**[2], **Youngmi Kim**[2], **Vikram Devgan**[2], **Patrick Danaher**[2], **Michael Rhodes**[2], **Mark Gregory**[2], **Sarah Warren**[2], **Joseph Beechem**[2]

1 Singapore Immunology Network (SIgN), Agency for Science, Technology and Research (A*STAR), Singapore, Singapore
2 NanoString® Technologies, Seattle WA

Microglia, the major tissue-resident macrophage in the parenchyma of the central nervous system, play key roles in brain development, homeostasis and neurological disorders. Dysregulation of microglia functions causes neurodevelopmental and neurodegenerative diseases. Therefore, a better understanding of microglial function and molecular crosstalk with brain cells is key to find new targets for disease treatment.

However, studying human microglial is challenging because (1) microglia tend to lose their unique gene expression patterns rapidly once isolated and (2) access to fresh human brain is very limited. For these reasons, most of microglial studies have used rodent brains as model systems despite significant gene expression profile differences compared to human brains. Thus, alternative approaches are required to overcome the drawbacks of rodent brain models. Recent advances in stem cell research have made it possible to generate three-dimensional (3D) brain organoids from human induced pluripotent stem cells (iPSCs) that have diverse neural cells and functions, recapitulating embryonic human brain development. In addition, single-cell RNA-sequencing (scRNA-seq) data have shown that gene expression profiles of brain organoids are similar to those of the human brain in the embryonic stages. However, brain organoids currently remain an imperfect model system since some key components of a brain, such as vascular endothelial cells and microglia, are missing in conventional brain organoids.

In an effort to construct a more complete brain organoid, we have established a method to generate human primitive macrophages (hiMacs) from iPSCs and succeeded in generating microglia-like cells (hiMicros) by co-culturing hiMacs in brain organoids. Our scRNA-seq data have shown that hiMicros regulate the differentiation of Neural Progenitor Cells (NPCs) within the brain organoids by modulating lipid metabolism, resulting in the reduction of NPC and neuron numbers as well as smaller organoids. To gain a deeper understanding of hiMicro biology in normal and disease contexts, we investigated the spatial location of hiMicro and their interaction with NPCs and neurons at the single-cell level using Spatial Molecular Imager (SMI) technology. SMI enabled us to visually examine spatial associations between hiMicros and NPCs and/or neurons in brain organoids with single cell resolution. It also enabled examination single cell gene expression profiles of these cell types in conjunction with their spatial relationships in developing brain organoids.

Multi-omic analysis of whole-transcriptome and high-plex protein assays on a single FFPE slide

Other Biology
Omics

POSTER **440**

Shilah A. Bonnett[1], **Nicholas Confuorto**[1], **Kit Fuhrman**[1], **Giang Ong**[1], **Alyssa Rosenbloom**[1], **John Lyssand**[1], **Gary Geiss**[1*], **Joseph Beechem**[1]

1 NanoString® Technologies, Seattle, WA
*Presenting author

The GeoMxTM Digital Spatial Profiler (DSP) enables high-plex, high-throughput spatial profiling and quantification from a single slide for either protein or RNA. To fully understand the interplay between RNA and protein in tissue, we have developed a novel dual analyte assay that allows for the profiling of both analytes from the area of interest (AOI) on a single slide. Here we describe the development and performance of the dual analyte assay on cell pellet array (CPA) and various tissues including tonsil and colorectal cancer (CRC) using a high plex GeoMx Protein Panel and the GeoMx Human Whole Transcriptome Atlas (WTA).

We successfully identified conditions to detect RNA and protein simultaneously, with comparable sensitivity and specificity to the single analyte conditions. The data accurately matched cell lines in an unbiased correlation analysis between GeoMx WTA and the entire CCLE RNAseq database. Furthermore, the assay performance was maintained in tissue samples. Tonsil and colorectal cancer biopsies were profiled for RNA (whole transcriptome) and protein from the same AOI, allowing simultaneous measurements of transcriptional and translational regulation of key biological pathways. This multi-analyte profiling enables deeper characterization of precious biological samples that are available in limited quantities.

Technology Development

Solid-phase DNA storage media for single-cell multi-omics with short- and long-read sequencing

Anna C. Hooker[1], **Sharmili Roy**[1], **Heon Seok Kim**[1], **Susan M. Grimes**[1], **Billy T. Lau1,**[2*], **Hanlee P. Ji**[1,2]

1 Division of Oncology, Stanford School of Medicine, Stanford CA
2 Stanford Genome Technology Center, Stanford University, Stanford CA
*Presenting author

Single-cell sequencing has enabled significantly deeper insight into contributions by individual cells in the broader context of biological diversity, tissue development and human phenotype. For example, single-cell RNA sequencing (scRNA-seq) and multi-omics methods that include spatial transcriptomics allows one to study the interactions among immune, stromal and cancer cells in primary tumors. However, major challenges for these types of multi-omic studies include (1) finite amounts of cells and single-cell or spatial libraries limiting follow-up genomic sequencing analysis; (2) innate complexities associated with multiplexed enrichment of targets of interest; and (3) computational difficulties arising during integration of multiomic data.

To address these changes, we designed a multiomic (single-cell and spatial) sequencing platform called CT-RAM (cellular transcript random access memory). This technology enables repeated and multiplexed interrogation of single cell transcriptomes and multi-omic targets without loss of original source analyte. Our technology leverages "click chemistry" to covalently attach sequencing libraries to either a bead- or tube-based solid-phase substrate; linkage occurs between functionalized tetrazine groups on the solid-phase substrate and trans-cyclooctene groups on the DNA molecules. By permanently conjugating libraries to a solid-phase substrate, the original cellular genomic library is preserved during iterative polymerase-based assays, allowing for repeated non-destructive molecular interrogation at single-cell resolution with spatial features. CT-RAM enables both analysis of specific gene targets and examination of full genomic features across different sequencing platforms. Moreover, polymerase-based multiplexed targeted enrichment of specific genes is expediated in our approach compared to conventional methods such as bait hybridization.

We utilized our novel technology platform to comprehensively assess mutations, gene expression, and isoform usage from single-cell transcriptome libraries. We performed multiple levels of omic analysis by gene-specific enrichment in a two-step nested extension and amplification reaction, followed by short-read (Illumina) sequencing for gene expression quantification, and long-read (Oxford Nanopore) sequencing for isoform and haplotype analysis.

As another usability demonstration of CT-RAM, we generated a spectrum of CRISPR edits to cells, and conjugated the resulting scRNA-seq libraries. Among individual cells, we directly confirmed CRISPR edits based on changes in specific target transcripts. Further, we achieved target enrichment averaging orders of magnitude compared to the baseline expression of each given gene target prior to amplification. Overall, CT-RAM enables a stable DNA storage system that allows for highly multiplexed targeted amplification and subsequent multilevel omic analysis from any cellular or molecular feature of interest.

Improved detection of unique transcripts from challenging samples using a novel, rapid-whole-transcriptome sequencing workflow

Technology Development

POSTER 501

Travis Sanders[1], **Lee French**[2], **Julie Walker**[1], **Clara Ross**[1], **Thomas Harrison**[1], **Ross Wadsworth**[2]

1 Watchmaker Genomics, Boulder, CO
2 Watchmaker Genomics, Cape Town, South Africa

RNA-seq is a powerful research tool that has increasing utility in clinical diagnostic applications, but there are many opportunities to improve RNA-seq workflows and performance. Many clinically relevant samples are not amenable to mRNA capture-based RNA-seq protocols. For example, template damage in FFPE-derived RNA samples results in significant 3'-bias, while data from overabundant globin mRNA in blood-derived samples are typically uninformative. Moreover, magnetic bead-based mRNA capture workflows are challenging to automate and prone to variability. Whole transcriptome sequencing (WTS), where rRNA and overabundant mRNA transcripts are depleted from the sample, accommodates these challenging sample types and offers a comprehensive view of the transcriptome, including biologically relevant non-coding transcripts.

The traditional RNA depletion approach employs DNA probes targeting RNAs to ensure robust depletion via RNase H. To minimize off-target effects, these protocols require a lengthy probe hybridization step before RNase H is added at an elevated temperature, which complicates the workflow and makes it difficult to automate. We hypothesized that factors such as probe design, duration of depletion, and high probe concentration may drive off-target depletion. To streamline the protocol while maximizing preservation of non-target RNAs, we developed algorithms to design optimized depletion probes. Using novel probe designs and chemistries, we systematically optimized reaction formulations and protocol parameters to support efficient depletion without the need for a separate probe hybridization step. This new substantially simplified workflow is easier to automate and reduces the total time required from 90 to 40 minutes.

To further improve the sample preparation workflow, we streamlined the cDNA synthesis and library construction process by reducing incubation times and eliminating cleanup steps, allowing conversion of samples into sequencer-ready libraries within 4.5 hours. We dramatically improved yields of low input samples by employing engineered enzymes and novel, optimized reaction formulations. Using RNA obtained from a range of sample types, including FFPE and blood, we compared our solution to commercial products. Spike-In RNA Variants (SIRVs) and ERCC RNA Spike-in Mixes were used to assess isoform detection and expression profiling, respectively. Libraries generated using our novel workflow resulted in less than 1% of reads mapping to rRNA and globin mRNA regions with no measurable off-target effects. We observed a high degree of variant calling concordance between independently characterized and annotated human reference and FFPE-derived samples. Expression profiling and isoform detection were robust across input amounts and technical replicates. As we anticipated, higher library yields translated to improved unique transcript identification.

High-sensitivity single-cell RNA profiling of fixed cells via in situ RNA-templated ligation

Technology Development

POSTER 502

Andrew Hill[1], **Jawad Aboudoud**[1], **Francis Cui**[1], **Andrew Kohlway**[1*], **Paul Lund**[1], **Nico Sponer**[1], **Joseph Arthur**[1], **Maria Alexis**[1], **Shaun Jackman**[1], **Shamoni Maheshwari**[1], **Sharmila Chatterjee Bhattacharjee**[1], **Kevin Chen**[1], **Winnia Hui**[1], **Martin Sauzade**[1], **Ryan Stott**[1], **Katherine Pfeiffer**[1]

1 10x Genomics, Inc., Pleasanton, CA
*Presenting author

High sensitivity single cell RNA profiling of fixed cells via in situ RNA-templated ligation Single cell RNA sequencing (scRNA-seq) has become a critical tool for profiling heterogenous mixtures of cells inherent to many biological systems. Increasingly scRNA-seq has been applied to profile tens to hundreds of samples corresponding to large cohorts of individuals or numerous perturbations, placing paramount importance on scalability and economy of sample collection, transport, storage and processing for scRNA-seq at scale.

Formaldehyde fixation is used extensively to preserve cells and tissue for transport, storage, and batch processing. However, the protein-RNA crosslinks, RNA strand breaks, and decrosslinking requirements make it challenging to obtain high quality scRNA-seq data from formaldehyde-fixed cells.

We have developed a high-throughput method for quantitatively profiling whole transcriptome gene expression in formaldehyde fixed and permeabilized single cells via hybridization of probes to RNA in situ. We designed a comprehensive whole transcriptome probe set with probe pairs that hybridize to transcripts from more than 18k genes. Hybridized probes are ligated, barcoded, and sequenced, allowing for highly sensitive detection of transcripts in cells fixed with 4% formaldehyde. It is also compatible with Feature Barcode technology for antibody-based cell surface protein detection.

Our probe panel design increased detection sensitivity for most genes via 3-fold tiling in conjunction with 1-fold tiling of a small number of genes with ubiquitously high expression. Nuisance genes like ribosomal proteins were excluded from the panel.

These design choices, in addition to higher rates of usable data, lead to a significant reduction in sequencing costs.

The probe set was validated by collecting matched data derived from both our on-market gene expression solution and our new method on matched mixtures of 16 sample types per species. We were able to select probe sets that produce highly concordant measures of gene expression across samples when compared to our existing products.

(continue to page 144)

Technology Development

POSTER 502

(continued from page 143)

We demonstrated the utility of this assay, along with short and long-term stability of fixed samples, on several sample types, including peripheral blood mononuclear cells, healthy and diseased tissue, and on stimulated cells over short time intervals. This new technology enables single cell RNA profiling in the context of large-scale studies or studies with distributed sample collection where robust sample stabilization is needed. We anticipate that this technology will enable broader adoption of single cell genomics in clinical and translational research settings and will enable critical biological insights.

Paired genotyping, scRNA-seq, and long-read sequencing define splicing aberrations in hematopoietic clonal outgrowths

Technology Development

POSTER 503

Federico Gaiti[1], **Allegra Hawkins**[1], **Paulina Chamely**[1], **Ariel Swett**[1], **Xiaoguang Dai**[2], **Lloyd Kluegel**[1], **Celine Chen**[1], **John Beaulaurier**[6], **Alex Drong**[2], **Scott Hickey**[6], **Tarek Mouhieddine**[3], **Ronan Chaligne**[4], **Sissel Juul**[2], **Eoghan Harrington**[2], **Justin Taylor**[5], **Irene Ghobrial**[3], **Omar Abdel-Wahab**[4], **Dan Landau**[1]

1 Weill Cornell Medicine, New York, NY
2 Oxford Nanopore Technologies, New York, NY
3 Dana-Farber Cancer Institute, Boston, MA
4 Memorial Sloan Kettering Cancer Center (MSKCC), New York, NY
5 University of Miami, Miami, FL
6 Oxford Nanopore Technologies, San Francisco, CA

Splicing factor mutations are recurrent genetic alterations in blood disorders, highlighting the importance of alternative splicing regulation in human hematopoiesis. However, the study of splice-altering mutations in humans has been hampered by the inability to distinguish mutant and wildtype cells in patient samples and the scarce coverage of short-read sequencing over splice junctions.

However, the study of splice-altering mutations in humans has been hampered by the inability to distinguish mutant and wildtype cells in patient samples and the scarce coverage of short-read sequencing over splice junctions. To overcome these limitations, we developed GoT-Splice by integrating Genotyping of Transcriptomes (GoT; Nam et al. 2019) with Nanopore long-read single-cell transcriptome profiling and CITE-seq. This allowed for the simultaneous single-cell profiling of protein and gene expression, somatic mutation status, and alternative splicing. Our method enabled the capture of higher number of junctions per cell and greater coverage uniformity vs. short-read sequencing.

We applied GoT-Splice to bone marrow progenitor cells from patients with myelodysplastic syndrome (MDS) to study how mutations in splicing factor 3B subunit 1 (SF3B1) corrupt human hematopoiesis. High-resolution mapping of SF3B1mut vs. SF3B1wt progenitors revealed an increasing fitness advantage of SF3B1mut cells with myeloid differentiation, resulting in an expansion of SF3B1mut erythroid progenitors (EPs). Accordingly, SF3B1mut EPs displayed higher protein expression of erythroid lineage markers and up-regulation of genes involved in cell cycle and checkpoint controls vs. SF3B1wt cells. While SF3B1mut cells showed the expected mis-splicing increase vs. SF3B1wt cells, they exhibited distinct cryptic 3' splice site (ss) usage as a function of hematopoietic progenitor cell identity, displaying stage-specific mis-splicing during EP differentiation.

Lastly, splicing factor mutations observed in CH subjects provide an opportunity to interrogate the downstream impact of SF3B1 mutations prior to development of disease. Like MDS, the application of GoT-splice to progenitor cells from SF3B1mut subjects with clonal hematopoiesis (CH) revealed

(continue to page 146)

Technology Development

POSTER 503

(continued from page 145)

increased mutant cell frequency in EPs and cell-type specific cryptic 3' ss usage in SF3B1mut cells. In summary, we developed a novel multi-omics single-cell toolkit to examine the impact of splicing factor mutations on cellular fitness and its molecular determinants from the earliest phases of clonal blood disorders (CH) to overt disease, providing a widely applicable method to study the effect of somatic mutations on isoform-level gene expression patterns in single cells directly in human samples.

Nucleus hashing applications with cell-type bias correction

Technology Development

POSTER 504

Isabelle Schmutz, Heather Geiger, Jade Carter, Salima Soualhi*

New York Genome Center, New York, NY
*Presenting author

Nucleus hashing is a multiplexing method for single nucleus RNA-seq (snRNA-seq) platforms that allows for uniquely labelling isolated nuclei, enabling pooling of samples for which single cell suspensions are difficult to obtain, such as complex tissues (e.g., brain) and archived frozen tissue (Gaublomme et al., 2019). Nucleus hashing uses oligo-conjugated antibodies against nuclear pore complex proteins to barcode nuclei of distinct samples. As tags are sequenced alongside the transcriptome and given a cell specific barcode, nuclei can be demultiplexed computationally, assigning them to a single sample of origin ("Singlet") or identifying them as inter-sample multiplets that are excluded from analysis.

Here we validated the reproducibility and robustness of nucleus hashing for snRNA-seq in the context of human postmortem brain tissue and mouse fresh frozen brain. We first set out to optimize our nuclei preparation protocol and compared three methods side by side: nuclei from crude homogenate, nuclei purified over 1.8M sucrose and nuclei purified over a 29% iodixanol density cushion. We found that centrifugation of nuclei over iodixanol efficiently eliminated myelin and reduced ambient RNA and mitochondrial RNA, yielding high quality nuclei suitable for single nucleus sequencing.

Hashing data for all three nuclei preparation protocols showed a relatively low signal to noise ratio of hashtag expression that is typical for nucleus hashing (Mylka et al., 2020), and resulted in a higher-than-expected number of nuclei for which sample origin could not be assigned ("Negatives"). To investigate the origin of the negatives we analyzed their RNA expression profiles alongside nuclei called as singlets in an experiment of hashed mouse cortex nuclei. We observed that while some of the negatives formed a distinct low-quality cluster in a UMAP projection, ~80% of the negatives clustered with nuclei identified as singlets, predominantly with oligodendrocytes. Similarly, in a hashing experiment of human brain samples, oligodendrocyte and astrocytes were overrepresented amongst the negatives. These data suggest that nucleus hashing and its demultiplexing can introduce a bias to recover neuronal cells over glial cells. Here we present our strategies to reduce the number of negatives and hence mitigate the cell type bias, both for the wet lab by reducing sample heterogeneity as well as computationally, through optimized analysis pipelines and the normalization of hashtag data. We are currently working on adapting nucleus hashing for the 10x Genomics multiome assay, to enable multiplexing and simultaneous detection of open chromatin and gene expression.

Gaublomme Jellert T. et al. 2019. "Nuclei Multiplexing with Barcoded Antibodies for Single-Nucleus Genomics." Nature Communications 10 (1): 2907.
Viacheslav Mylka et al. 2020. "Comparative analysis of antibody- and lipid-based multiplexing methods for single-cell RNA-seq." Biorxiv, doi: https://doi.org/10.1101/2020.11.16.384222

Automated processing of solid tissues into single cells or nuclei for genomics and cell biology applications with the Singulator™ 100 and 200 systems

Technology Development

POSTER 505

Stevan Jovanovich, John Bashkin, Kaiwan Chear, Nabiha Khan, Kaushal Kumbagowdana, Bruce Leisz, Roger McIntosh, Danielle Meyer, Nathan Pereira

S2 Genomics, Livermore, CA

Single-cell sequencing is revealing the next level of complexity of biological systems. The ability to measure the heterogeneity of single-cell gene expression and chromatin structure cell by cell has led to the discovery of new cell types and expanded our understanding of the diverse regulatory pathways and states of individual cells. Single-cell sequencing workflows for multiple applications require reproducible generation of high-quality single-cell or nuclei suspensions.

S2 Genomics has developed and commercialized the patented Singulator™ 100 System to automate the dissociation of solid tissues into single cell or nuclei suspensions. The Singulator disrupts solid tissues in single-use cartridges using coupled enzymatic/chemical dissociation and mechanical disruption.

S2 Genomics has now developed the Singulator 200 system, with higher sample throughput, improved temperature control, and other advancements. The operation of the Singulator 100 and 200 systems for the dissociation of mouse and human solid tissues will be described, along with new, advanced software for protocol customization and improved processing of samples <10 mg. Validation data will be presented for processing <5 to 400 mg samples of fresh mouse and human solid tissues, including matched sets of normal and cancerous patient samples, into single-cell suspensions with high viabilities and yields. In addition, data of processing fresh or frozen mouse and human tissue samples into high-quality nuclei in <6 min will be presented.

The Singulator-prepared samples have been analyzed by single-cell RNA sequencing (scRNA-seq), single-nucleus RNA sequencing (snRNA-seq), and assay for transposase-accessible chromatin using sequencing (ATAC-seq) with a nanodroplet single-cell library preparation system. The best methods for downstream processing, including additional sample purification and use of RNase inhibitors will be discussed. In addition, data will be presented on the use of the Singulator systems to isolate specific cell types from tumor samples will be shown.

Supported in part by National Institutes of Health award number R41 HG010129-01 to Dr. Jovanovich.

Maximizing the output of RNA-Seq for low-input, challenging samples using the xGen™ Broad Range RNA Library Prep Kit

Technology Development

POSTER 506

Bita Lahann[1], Sukhinder Sandhu[1], Laurie Kurihara[1]

1 Integrated DNA Technologies, Coralville, IA

Most labs consider the depletion of ribosomal RNA (rRNA) (ribodepletion) as the default approach for processing formalin fixation and paraffin-embedding (FFPE) RNA samples – despite high costs for reagents and sequencing – since poly(A)-selection would only provide 3' coverage of fragmented transcripts. However, hybridization capture for enriching desired targets can also be used to produce quality data from FFPE RNA-Seq libraries. Commonly utilized commercial workflows for RNA-Seq library preparation are vulnerable to crosslinked or fragmented RNA substrates as well as an increased representation of intronic reads in the final library, an observation that is consistently seen for formalin-fixed samples. In addition, libraries produced from limited material suffer from high duplication rates and adapter dimers, as well as lower mapping rates and fewer genes and transcripts detected.

As a solution for FFPE samples in research, we present the IDT xGen Broad Range RNA Library Prep Kit. This workflow leverages patented Adaptase™ technology, a method for adapter ligation to single-stranded substrates, to directly convert first-strand cDNA into NGS libraries in 4.5 hours. This approach has been shown to save time, stranded libraries without requiring conventional second-strand cDNA synthesis or degradation and increases yield and complexity while reducing bias and undesired library. Here, we present workflow options and summarize the resulting tradeoffs to data and costs. A price per sample savings advantage is presented through the hybridization capture workflow relative to ribodepletion when exonic transcript information is the desired data output. Due to the high variability in quality and quantity of RNA derived from FFPE, we present data to support the use of the IDT xGen Broad Range RNA Library Kit with challenging and low input RNA samples.

*For research use only. Not for use in diagnostic procedures. Unless otherwise agreed to in writing, IDT does not intend these products to be used in clinical applications and does not warrant their fitness or suitability for any clinical diagnostic use. Purchaser is solely responsible for all decisions regarding the use of these products and any associated regulatory or legal obligations.

Successful application of mostly natural sequencing by synthesis to single-cell RNA-seq

Technology Development

POSTER 507

Joshua Z. Levin[1,2], **Sean K. Simmons**[1,2*], **Gila Lithwick-Yanai**[3*†], **Xian Adiconis**[1,2], **Florian Oberstrass**[3], **Nika Iremadze**[3], **Kathryn Geiger-Schuller**[1,4], **Pratiksha I. Thakore**[1,4], **Omer Barad**[3], **Orit Rozenblatt-Rosen**[1,4], **Aviv Regev**[1,4,5], **Doron Lipson**[3]

1 Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA
2 Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA
3 Ultima Genomics, Newark, CA
4 Genentech, South San Francisco, CA
5 Howard Hughes Medical Institute and Koch Institute of Integrative Cancer Research, Department of Biology, MIT, Cambridge, MA
*Authors contributed equally
†Presenting author

The emergence of next generation sequencing has had a revolutionary impact in both basic science and medicine. One of the key factors driving this has been the ability to generate large-scale datasets at lower costs. Here, we describe a new sequencing technology that has the potential to advance the field in this direction. Mostly-natural Sequencing by Synthesis (SBS) leverages a low fraction of labeled nucleotides to combine the speed and read length advantages of non-terminating chemistry with the throughput and scalability of optical endpoint scanning to enable low-cost, high-throughput sequencing. We decided to test this new type of sequencing with single-cell RNA sequencing (scRNA-seq) because scRNA-seq, which has become a central tool for identifying and characterizing cell types, states, and circuitry, often requires large-scale sequencing.

To benchmark mostly-natural SBS with scRNA-seq, we performed experiments with three sample types: (1) PBMCs from a single individual, (2) PBMCs from a mixture of eight individuals, and (3) Perturb-seq. In the first experiment, we compared the results of sequencing the 5' or 3' ends of the transcript. Since the mostly-natural SBS generates single end reads, the 3' end reads require reading through a poly(T) sequence. Because lower cost sequencing methods would also enable large-scale Perturb-seq screens with their heavy sequencing requirement, we tested mostly-natural SBS with a large Perturb-seq screen in human A375 cells targeting 1902 transcription factors and chromatin modifiers. For each experiment, we also performed SBS (Illumina) as a control for comparative purposes.

In general, we found that the results were similar between mostly-natural SBS and SBS. Some of the minor differences that we observed seem to be due to single end reads with mostly-natural SBS vs. paired end reads with SBS. We will discuss computational adjustments that can help ameliorate issues resulting from single end reads. We will present comparisons of metrics for both basic performance, such as the alignment of reads and sensitivity, and for their ability to recover known biological information. Finally, we will examine the impact of increasing sequence depth in the context of scRNA-seq.

One-step reverse-transcription and variant-avoidant multiplex targeted amplification of SARS-CoV-2 for whole-genome sequencing

Technology Development

POSTER 508

Kaylinnette Pinet[1], **Bradley W. Langhorst**[1], **Yanxia Bei**[1], **Lynne Apone**[1], **Luo Sun**[1], **Pingfang Liu**[1], **Andrew Gray**[1], **Julie Menin**[1], **Katherine Marks**[1], **Kerri Sakowski**[1], **Nicole M. Nichols**[1], **Fiona J. Stewart**[1], **Eileen T. Dimalanta**[1*], **Theodore B. Davis**[1]

1 New England Biolabs, Ipswich MA
*Presenting author

The increasing prevalence of SARS-CoV-2 variants is an on-going global issue with variants of concern accounting for more than 90 percent of recent SARS-CoV-2 sequence submissions to GISAID from over 30 nations. A standard approach for SARS-CoV-2 whole-genome sequencing requires cDNA synthesis, followed by targeted multiplex PCR amplification, then library preparation and sequencing. However, the accurate sequencing of current and emergent variants requires faster and more resilient methodologies. We have streamlined this sequencing approach by combining the reverse transcription and targeted amplification steps into a single step through the application of a reverse transcriptase and high-fidelity DNA polymerase enzyme mix. Combining the reverse transcription and targeted amplification reactions has the added benefits of reducing plastic consumption and probability of sample cross-contamination.

We further improved this sequencing approach by designing novel primer mixes for targeted amplification of SARS-CoV-2 that reduce the impact of variants on amplification efficiency. Early primer schemes for SARS-CoV-2 targeted amplification (i.e., ARTIC v1-3 and Midnight primers) were designed in 2020 based on the original genome sequence derived early in the pandemic. When using ARTIC v1-3 or Midnight primers to sequence current SARS-CoV-2 samples, some amplicons may show reduced representation due to the presence of variants at priming sites. To address this issue, we utilized over a million SARS-CoV-2 GISAID sequences uploaded by March of 2021 to design variant avoiding primers.

These SARS-CoV-2 variant-avoidant primers generate seventy-four ~550 bp amplicons from standard and variant SARS-CoV-2 templates and they can be employed upstream of Oxford Nanopore Technologies® or Illumina® sequencing platforms. They also utilize the same overlapping amplicons and dual primer pool approach as ARTIC and Midnight-based SARS-CoV-2 sequencing methodologies. Therefore, these innovative primer mixes are compatible with the majority of existing ARTIC or Midnight-based sequencing analysis pipelines. These primers in conjunction with the single-step cDNA synthesis and multiplex targeted amplification protocol can help support current and future global efforts to track and monitor variants of concern.

High-throughput plasmid sequencing transforms the scale of synthetic biology screens and accelerates cell therapy development

Technology Development

POSTER 509

Sofia Kyriazopoulou Panagiotopoulou, Howard Kim, Li Wang, Emily Damato, Sergey Timokhin, Aaron Cooper, Andrea Chan, Ryan Fong, John Gagnon, Brendan Galvin, Fabio Ingrao, Meng Lim, Marian Sandoval, Cate Sue, Anzhi Yao, Grace Zheng, Tarjei Mikkelsen

Arsenal Biosciences, South San Francisco, CA

Developing cell therapies using complex synthetic biology requires the exploration of vast design spaces. Yet, the time and cost associated with the design, construction, and quality control of large-scale screening libraries have limited the applications and scope of past efforts. To overcome these challenges, we developed SKIP (Sequence Knowledge of Inventoried Plasmids), an NGS-based workflow and computational pipeline for rapid, cost-efficient, and highly scalable verification of large plasmid libraries. We have utilized SKIP to power synthetic biology screens in human primary T cells for the purpose of developing safe and efficacious cell therapies against cancer.

Advances in cloning techniques have made it possible to synthesize large libraries with thousands of unique, multi-kilobase members in a pooled reaction, combining fragments randomly to achieve combinatorial diversity in design space. Standard sequence verification by Sanger sequencing is low-throughput and costly, and requires painstaking manual analysis, thus creating a bottleneck downstream of these new build capabilities. Recently described approaches for short-read sequencing of plasmids address some of these limitations, but have complicated workflows and often produce data that suffer from uneven sequencing coverage that hinders efficient computational analysis.

Our in-house developed pipeline, SKIP, can perform NGS-based, full-length sequence verification of plasmids at scale. This semi-automated workflow, based on bead-linked transposomes, enables short-read sequencing of complete plasmids directly from bacterial clones, without the need for plasmid purification or input plasmid normalization. Furthermore, it can operate with minuscule amounts of input plasmid material. The NGS reads generated by SKIP have minimal sequence biases, uniform coverage, and enable de novo assembly of plasmid sizes exceeding 10kb. A scalable computational pipeline can verify clones and accurately match them to unique members of a combinatorial library design through de novo assembly and variant calling of thousands of sequencing libraries in a matter of hours.

We have applied the highly efficient SKIP workflow to our ICT (Integrated Circuit T Cell) development process, to assemble, sequence-verify, and screen thousands of CAR (chimeric antigen receptor) and shRNA payloads for functional attributes associated with increased potency and safety.

Development of a novel qPCR assay to detect SARS-CoV-2 RNA extracted from wastewater

Technology Development

POSTER 510

Mark Sharkey[1], Kristina Babler[2], George Grills[3], Mario Stevenson[1], Helena Solo-Gabriele[2]

1 Department of Medicine, University of Miami Miller School of Medicine, Miami, FL
2 Department of Civil, Architectural, and Environmental Engineering, University of Miami, Coral Gables, FL
3 Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, FL

Since the onset of the COVID-19 pandemic, sensitive and accurate diagnostic assays have been essential to detect infected individuals and support efforts to limit further spread of the virus. We initially focused on developing a simple, low-cost assay ($1US) that could be used to directly detect the RNA of SARS-CoV-2 (the virus that causes COVID-19) in unprocessed human saliva, and then applied this assay for detection in wastewater. The assay utilizes PCR primers and a reporter probe designed to target the SARS-CoV-2 nucleocapsid gene, and a novel polymerase called Volcano 2G (V2G) that efficiently reads both RNA and DNA templates. The V2G-qPCR assay is sensitive and specific, simple to execute on standard qPCR instruments, and provides an instantaneous visual readout at PCR endpoint through blue light excitation.

Minor modifications to the standard assay allowed for quantitative PCR using a real-time qPCR instrument. Nucleocapsid DNA standards were synthesized by PCR, quantified by droplet digital PCR and used to generate standard curves for determining SARS-CoV-2 RNA quantities of unknowns by qPCR. Based on the standard curves, amplification efficiency is greater than 95%, with a coefficient of determination of 1, or very close to 1. V2G-qPCR assays have also been developed to measure targets to control for sample processing (OC43) and to measure the degree of PCR inhibition (HIV RNA). Assays to measure PMMoV and human beta-2-microglobulin RNA have also been developed for standardization of CoV-2 RNA copy data.

For weekly sampling, wastewater was concentrated on electronegative filters, and total RNA was purified manually using silica based spin columns. Importantly, eluting the RNA with a minimal volume and increasing the wash buffer volumes was effective in reducing PCR inhibition to negligible levels. SARS-CoV-2 RNA levels, as determined by V2G-qPCR ranged from 102 to 106 genomic copies per liter raw wastewater and correlated well with quantitation by a more commonly used commercially available RT-qPCR assay. Although standard RT-qPCR assays perform similarly, notably RNA extracted from wastewater is amplified directly by V2G-qPCR, and thus the cDNA synthesis step is bypassed, thereby reducing both assay time and cost compared to standard RT-qPCR. Overall, V2G-qPCR is an accurate and cost-effective approach to directly measure target RNA in less than one hour.

This work was supported financially by the University of Miami (Coral Gables, FL) administration and the National Institute on Drug Abuse of the National Institutes of Health (NIH) under Award Number U01DA053941.

Robust and highly efficient RNA-seq library preparation workflow for new sequencing platforms

Technology Development

POSTER 511

Keerthana Krishnan, Chen Song, Matthew Campbell, Pingfang Liu, Bradley W. Langhorst, Eileen T. Dimalanta, Theodore B. Davis

New England Biolabs, Ipswich, MA

Next Generation Sequencing has revolutionized the way we approach scientific questions and RNA-seq (RNA sequencing) has undoubtedly become the most popular method for transcriptome analysis. As RNA-seq is being increasingly adopted for molecular diagnostics, quality and reproducibility of library preparation methods are becoming more important. Concurrently, to meet the unprecedented demand and need for increased throughput, better quality, cheaper workflow options, advances are being made in sequencing instruments too. New competitive chemistries eg. MGI are outperforming market leaders in terms of cost, speed and amount of data being generated.

In this study we use a new and improved RNA-seq workflow across sequencing instruments to generate high quality data from multiple RNA inputs. We analyze gene body coverage, transcript abundance, GC content and duplication to show superior data on the new MGI sequencing platform. As a result, our method has increased sensitivity and specificity, delivering high quality strand-specific data. Our method has accompanying modules for polyA tail enrichment and Ribodepletion and is amenable to large-scale library construction and automation.

Adapting Enzymatic Methyl-seq (EM-seq) for long-read sequencing

Technology Development

POSTER 512

Paul W. Hook[1], Felix Krueger[2], Winston Timp[1]

1 Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD
2 Bioinformatics, The Babraham Institute, Cambridge, United Kingdom

The accepted gold standard for measuring DNA methylation is whole genome bisulfite sequencing (WGBS). Bisulfite sequencing works through conversion of unmethylated cytosines to thymines through chemical treatment and PCR, usually followed by conventional Illumina sequencing. This approach has been largely limited to short reads because bisulfite treatment is harsh and results in DNA fragmentation. This prevents phasing methylation status between distal CpG sites, obscuring potentially important epigenetic coordination. Long read technologies have overcome this limitation by measuring native methylation, allowing for phasing across the genome. However, the amount of unamplified DNA needed for long read approaches is unattainable for many biological applications, including clinical samples. Here, we demonstrate an implementation of Enzymatic Methyl-seq (EM-seq) (NEB) for long read methylation sequencing. Instead of chemical treatment, EM-seq uses enzymatic treatment followed by PCR amplification. This approach leaves DNA molecules largely intact while enabling an amplification step.

We optimized our approach by characterizing genome-wide CpG methylation in the GM12878 cell line with Nanopore and PacBio long read sequencing. This optimization included modifying handling procedures in order to retain long molecules, testing different polymerases during amplification, tuning the number of amplification cycles, and increasing the elongation times during amplification. With this data, we modified the well-established and popular methylation calling software, Bismark, to use the long read aligner minimap2. By tuning the performance of long read Bismark (lrBismark) for aligning long reads to a three-base genome, we were able to increase speed by greater than 100-fold over default settings. With this data, we found we can consistently measure CpG methylation on long reads ranging between 1-7kb with a starting input between 200-500ng. We then combined these new methods with targeted capture (Twist Bioscience) to measure CpG methylation in promoters of protein-coding genes in the human genome.

By optimizing both molecular and computational approaches, we have implemented a robust protocol for long read (up to ~7kb) methylation profiling on limited material. This combination of new molecular approaches here sets the stage for bespoke assays that measure CpG methylation in areas of interest from low input samples with long reads. Further, the modification of Bismark provides a readily accessible and familiar data analysis tool, allowing these new approaches to be easily integrated into downstream analyses.

High-throughput RNA isoform sequencing via programmable cDNA concatenation

Technology Development

POSTER 513

Aziz Alkhafaji

Broad Institute of MIT and Harvard, Cambridge, MA

While RNA-sequencing has dramatically accelerated our understanding of biology, quantitation and discovery of full-length RNA isoforms resulting from alternative splicing remain poorly resolved. Alternative splicing is a core regulatory process that modulates the structure, expression, and localization of expressed proteins through differential exon and/or UTR splicing during transcript maturation. Beyond being an integral component of cellular development and homeostatic maintenance, RNA splicing is implicated in a wide range of pathologies with hallmark isoforms being linked to cardiovascular, neurological, and immunological diseases. Current limitations in isoform quantitation and discovery arise from the inability of existing sequencing platforms to scalably sequence full-length mRNAs – short-read approaches are unable to span most successive splice sites while long-read approaches are constrained by limited throughput.

To enable high-throughput full-length RNA sequencing, we have developed the method Multiplexed Arrays sequencing (MAS-seq) which maximizes the sequencing potential of isoforms on the PacBio platform. Through the use of deoxy-uracil digestion followed by barcode-directed ligation of cDNAs, MAS-seq generates long multiplexed cDNA arrays with a length distribution that allows for both accurate consensus sequencing and optimal capacity utilization of PacBio sequencers. In combination with upstream artifact depletion measures, MAS-seq boosts the sequencing throughput to approximately 40 million full-length transcripts per SMRT Cell 8M flow cell, a >20-fold increase compared to PacBio's scIso-Seq workflow. We sequenced synthetic RNA isoform standards (Lexogen) and demonstrate a 99.8% isoform identification accuracy using MAS-seq, far exceeding the 56.8% accuracy of Smart-seq3, the best-in-class short-read isoform identification protocol. MAS-seq-enabled single-cell RNA isoform sequencing of tumor infiltrating CD8+ T cells robustly identifies the canonical CD45 (PTPRC) isoforms associated with the range of observed T cell states - a finding additionally validated at the protein level via CITE-seq. Further, we demonstrate the impact of the long-read sequencing throughput gains on single-cell isoform analysis enabled by MAS-seq, providing a 44% increase in single-cell clustering capacity (adjusted Rand index) and a 34-fold gain in identifying differentially spliced genes amongst the CD8+ T cell subtypes. MAS-seq is a streamlined and cost-effective approach that enables high-throughput bulk and single-cell RNA isoform sequencing.

Quantification of gene expression changes in mouse disease models using a high-throughput spatial omics platform

Technology Development

POSTER 515

Tarif Awad, Brett Cook, Natalya Kotova, Irene Oh, Marcos Otero-Garcia

Rebus Biosystems Inc., Santa Clara, CA

Batch effects due to technical variability are a major problem in single cell transcriptomics. Spatial methods are no exception – their low throughput requires high numbers of technical replicates, reducing the statistical power needed to quantify differential gene expression across experimental conditions. To overcome this problem, we took advantage of the large imaging area of the Rebus Esper spatial omics platform. We processed brain sections from three mouse genotypes in parallel - one wild type and two disease models. In addition, the Esper High Fidelity assay, based on single-molecule fluorescent in situ hybridization (smFISH), requires no amplification, yielding quantitative results with minimal batch effects. The combination of low technical variation and balanced experimental design allowed us to integrate more than 500,000 cells from multiple datasets for analysis without the need for batch correction. We were able to identify more than 12 neuronal and glial cell type clusters using 20 cell type-specific genes, and further dissect these cell types by anatomical structures utilizing the spatial information. We then performed differential gene expression of 10 disease-related genes on each cell-type subset. These results demonstrate the ability of the Rebus Esper spatial omics platform to yield high-throughput spatial omics data with single cell resolution, low technical variation, and the sensitivity and specificity required for differential gene expression quantification.

Dynamically visualizing the link between histological structures and whole-transcriptome expression data to explore the Spatial Organ Atlas

Technology Development

POSTER 516

Megan Grout-Vandenberg[1], **Natalie Holland**[1], **John Thomas Hoffer**[2], **Jeremy L. Muhlich**[2], **Jason Reeves**[1], **Yan Liang**[1], **Alison Vanschoiack**[1], **Jingjing Gong**[1], **Erin Piazza**[1], **Kit Fuhrman**[1], **Patrick Danaher**[1], **Margaret Hoang**[1], **Sarah Warren**[1], **Nicholas Confuorto**[1], **Kit Fuhrman**[1], **Michael Rhodes**[1], **Joseph Beechem**[1]

1 NanoString Technologies, Inc., Seattle, WA
2 Laboratory of Systems Pharmacology, Harvard Medical School, Boston, MA

We adapted the Minerva suite to use data from NanoString®'s Spatial Organ Atlas, which includes both immunohistochemistry (IHC) images as well as Whole Transcriptome Atlas expression analyses in pancreas, kidney, colon, lymph node, and brain. The Minerva Author program inputs an OME-TIFF, a metadata-rich, open-source imaging filetype. We extracted region of interest (ROI) content from OME-TIFFs to overlay ROI masks over the IHC image, emphasizing the pathology-driven selection. ROI masks were color-coded by histological subtype to typify categories without need for extensive labeling. We contributed to the software code base to make masks transparent, so that the ROI selections were abundantly clear throughout the guide. Minerva Author allows the user to craft a "Story" of narrated waypoints, incorporating interactive plots and defined display settings. We added clickable SVG regions on embedded reference schematics of organs to unite the tissue sample IHC image with gross anatomy. Consequently, the user can click on a data point in a graphic, scatterplot, barplot, or heatmap and be taken instantly to the appropriate region of the tissue. SVG use was extended further to be supported in Story creation. Existing interactive graph functionality was also extended through SVGs. Built-in interactive visualizations, such as single-cell scatter plots, adjust display settings on the image in accordance with where the user clicks on the plots. In each case, the significance of the biological result was indicated with a change invoked in the IHC image: panning to, zooming in on, or applying corresponding transparent mask over ROIs of a histological subtype. Changes in Minerva software facilitate communicating Spatial Organ Atlas images, data, and experimental design in an accessible, interactive, and powerful way.

Enzymatic DNA synthesis (EDS) allows rapid access to synthetic oligos to produce expression vectors for SARS-CoV-2 spike variants

Technology Development

POSTER 517

Benoit Derrien[1], **Nelly Lourenco**[1], **Pablo Guardado-Calvo**[2], **Marija Backovic**[2], **Nadège Tardieu**[1], **Henri Lachaize**[1], **Maryke Appel**[1], **Christine Peponnet**[1], **Félix A. Rey**[2], **Xavier Godron**[1]

1 DNA Script, Le Kremlin-Bicêtre, France
2 UMR 3569, CNRS Virology Lab, Paris, France

The emergence of SARS-CoV-2 variants has created a significant demand for research materials to study the biology of these mutant viruses.

Recombinant proteins corresponding to the SARS-CoV-2 Spike (S) protein variants (S is the primary viral antigen) are required to study virus transmissibility and immune escape. Expression plasmids for structural and functional studies may be obtained via a variety of strategies, ranging from the assembly and cloning of PCR products to the production of full-length synthetic constructs. Whilst the latter appears to be the most advantageous approach, it is expensive and requires long turnaround times from a handful of commercial suppliers.

We have developed a novel enzymatic DNA synthesis (EDS) technology and a benchtop instrument that enables decentralized, same-day, on-demand oligo synthesis in a standard molecular biology lab. Here we report on the use of EDS primers for multi-site directed mutagenesis to generate variant versions of SARS-CoV-2 spike (S) protein, as well as sequence confirmation of recombinant plasmids by Sanger sequencing.

Mutagenesis primers were used to produce expression plasmids encoding SARS-CoV-2 UK/B.1.1.7 or Indian/B.1.617.1 spike protein variants for functional studies. Oligos were designed to introduce up to ten mutations into cloned copies of the 3.8-kb S gene using a multi-site approach. Sanger sequencing primers produced by EDS were used to either confirm individual mutations in picked clones, or for tiled sequencing of the entire S coding sequence (CDS). EDS primers enabled high success rates in multi-site mutagenesis, and yielded high-quality Sanger sequencing data.

Access to on-demand oligo synthesis for site-directed mutagenesis allowed us to develop and successfully employ a fast and flexible approach for the generation of SARS-CoV-2 S expression constructs as new variants emerge or study objectives evolve. Modification of existing plasmids is an iterative process that requires only a few days per cycle of mutagenesis and sequence confirmation, as compared to weeks of waiting time for full-length synthetic constructs.

Keywords: Enzymatic DNA Synthesis, SARS-CoV-2, site-directed mutagenesis, variant confirmation

Multiplexed single-cell and spatial transcriptomics profiling of immune cell markers in FFPE tumor tissues using the novel RNAscope™ HiPlex v2 in situ hybridization technology

Technology Development

POSTER 518

Sayantani Basak[1], **Michaeline Bunting**[1*], **Ming Yu**[1], **HaYeun Ji**[1], **Ching-Wei Chang**[1], **Bingqing Zhang**[1], **Anushka Dikshit**[1]

1 Advanced Cell Diagnostics, a Bio-Techne brand, Newark, C
*Presenting author

The tumor microenvironment (TME) is characterized by profound cellular heterogeneity and complex cell to cell interactions between tumor cells, immune cells, stromal cells, and extracellular matrix. Defining the spatial interactions between various cell types and their activation states is crucial for implementing successful immunotherapy strategies against various types of cancer. We utilized a highly sensitive and specific multiplexed technique, the RNAscope HiPlex v2 in situ hybridization (ISH) assay for spatial transcriptomic profiling of target genes to assess immune regulation in human lung, breast, cervical and ovarian FFPE tumor tissues at single cell resolution.

We have expanded our current RNAscope HiPlex assay capability of iteratively multiplexing up to 12 targets in fixed and fresh frozen samples to include FFPE tissues. A novel FFPE reagent in the assay effectively reduces background autofluorescence, improving the signal to noise ratio. We have leveraged this technology to investigate spatial expression of 12 oncology and immuno-oncology target genes, including tumor markers, immune checkpoint markers, immunosuppression markers, immune cell markers and secreted chemokine RNA expression profile within the TME. The targets were simultaneously registered using HiPlex v2 image registration software that enables background subtraction.

Using this assay, we visualized T cell infiltration and identified T cell subsets within tumors using CD3 and CD8 expression and activated T cells by IFNG expression. We further identified subsets of pro- and anti-inflammatory macrophages by CD68 and CD163 expression as well effector cells which secrete chemokines and cytokine. We also detected the hypoxia markers HIF1A and VEGF to elucidate the immunosuppressive state of tumor cells. Preliminary analysis and quantification of the HIF1A expression using HALO® image analysis software showed higher copy numbers in the lung tumor as compared to the other tumors, demonstrating the sensitivity of the assay through differential expression. We additionally showed the differential expression of immune checkpoint markers PDCD1, and CD274 within the TME.

In summary, using a highly sensitive multiplexed RNAscope HiPlex v2 ISH assay, we have demonstrated the capability of this technique to spatially resolve 12 targets in four different tumor types. Quantification of immune markers further depicted a differential expression among various tumors. This technology is highly beneficial for investigating complex and spatial tumor-stroma interactions and can provide valuable understanding of the biological interactions among various cell types in complex and heterogeneous tissues.

High-throughput molecular-weight DNA extraction, library preparation, and size selection for human whole genome sequencing

Technology Development

POSTER 519

Jeffrey Burke[1], **Renee Fedak**[1], **Duncan Kilburn**[1], **Michelle Kim**[1], **Heather Ferrao**[2], **Suzanne Dee**[2], **Kelvin J. Liu**[1]

1 Pacific Biosciences, Inc. (PacBio), Baltimore, MD
2 Pacific Biosciences, Inc. (PacBio), Menlo Park, CA

Rapid advances in long-read sequencing technology have led to dramatic improvements in cost and throughput. De novo sequencing projects using long-read sequencing to generate assemblies have grown from creating single reference genomes to examining population wide diversity across progressively larger groups of individuals, paving the way for a future of clinical human whole genome sequencing. Because of its high accuracy, Pacific Biosciences HiFi sequencing is already being used to study rare genetic conditions in pediatric cancer and to generate a human pan genome reference.

However, high throughput sample and library preparation methods are needed to support the generation of such large numbers of genomes. Such workflows are commonplace in short read sequencing but do not widely exist for long-read sequencing which requires manipulation of high molecular weight (HMW) DNA.

We present a high-throughput HMW DNA extraction, library preparation, and size selection workflow capable of processing 96 samples in as little as 8 hours for HiFi sequencing on Pacific Biosciences Sequel IIe systems.

Automated HMW DNA extraction is first performed using Nanobind magnetic disk technology on Thermo Fisher KingFisher instruments. In contrast to magnetic beads, each extraction uses a single Nanobind disk that is covered with micro- and nanostructured silica wrinkles to protect the bound DNA from shearing during automated processing. The method is capable of HMW DNA extraction from 96 cell, blood, and tissue samples in 60 – 180 minutes. Fully walkaway solutions are also available for use with Hamilton NIMBUS instruments.

Automated library preparation is then performed using PacBio Express Template 2.0 Kit on Hamilton NGS Star.

Finally, high throughput size selection is performed using a 96 well plate version of Short Read Eliminator technology. Short Read Eliminator technology uses a proprietary size selective precipitation chemistry to deplete small DNA from HMW DNA samples. Alternatively, a custom magnetic bead based size selection can be used to achieve high size selection cutoffs as well.

A variety of sample types, including cultured cells, bacteria, whole blood, and tissue are demonstrated. As these methods require only common commercial instrumentation, it can be easily integrated into many existing laboratory environments.

Integrating long-read and single-cell sequencing to improve biopharmaceutical production

Technology Development

POSTER 520

Kevin McFarland, **Natalia Majewska, Siqi Chen, Michael Betenbaugh, Winston Timp**

Johns Hopkins University, Baltimore MD

Population-level trends in clonal cell lines, such as loss of productivity, can be due to changes at the single cell level. In particular, the epigenome and transcriptome can be significantly heterogeneous even within a phenotypically homogenous cell population. Single-cell sequencing methods have been developed over the last decade and applied to cancer and development, but are underutilized in the biomanufacturing industry where consistency is key.
It is our hypothesis that there are subpopulations in the ostensibly homogenous samples with differing gene target productivity.

Here we focus on two Chinese hamster ovary (CHO) clonal cell lines which produce IgG. The clones were grown either with or without selection pressure (glutamine supplementation) for 30 cell doublings, harvesting cell pellets every 10 doublings. At each of these time points, we generated both whole-genome nanopore sequencing data and single-cell RNA sequencing data to probe shifts in the genome, methylome and single-cell transcriptome to better understand differences that arise and are then selected for in the population. Over this time period both clones dropped significantly in gene target productivity, accompanied by shifts in global methylation and changes in methylation in and around the gene target itself. We also observed copy number variation which arose in the notoriously unstable CHO genome over time. From the single cell data we identified different subpopulations arising in the (originally) clonal population over time which we attribute to differences in productivity by clustering based on expression of the human IgG gene target. We explored differences in the subpopulations comparing cell populations under selection or not under selection.

By integrating long read epigenetic and genetic profiling with single cell transcriptome sequencing, we can observe changes in cell populations over time and how selection pressure acts to influence CHO cell performance. This data will allow us to target these changes for cell engineering or bioprocess optimization to sustain consistent CHO growth and productivity metrics.

Human DNA enrichment for enhanced forensic identification

Technology Development

POSTER 521

Shawn E. Levy[1,2], **Nripesh Prasad**[2], **Braden Boone**[2], **Arianna Pionzio**[2]

1 HudsonAlpha Institute for Biotechnology, Huntsville, AL
2 Discovery Life Sciences, LLC, Huntsville, AL

Over the last decade, advancements in genomic technology and methodology, particularly in the sequencing space, have revolutionized the ability of scientists to profile a wide diversity of sample types and applications, including in the forensic sciences. When combined with data from widespread use of direct to consumer genealogy products in public databases, many identifications and prosecutions have occurred in otherwise cold cases due to genomic technologies. Our group has been fortunate to be involved in several unique cases and by advancing the methods to use less and less DNA from more challenging sources, we've continued to advance the possible in forensic identifications. We will present work that shows as little as 30pg of human genomic DNA can be successfully analyzed to a prosecution-quality result even in the presence of high amounts of non-human DNA. Optimizing from methods developed for analysis of environmental, oncology and rare cell sample, we have developed methods and analyzed over 800 forensic cases. Samples from charred bone, crime scene material, and trace evidence have all been successful. A recent case that involved trace amounts of semen has opened the door to solving a nearly 50 year old murder case and provided profiles to allow law enforcement to open additional cases. Combining highly sensitive library preparation methods with human genome enrichment has allowed trace DNA samples to be accurately examined and unambiguously identified. Updates to validation, accuracy and quality parameters will be discussed to help refine standards and metrics for continued advancement in forensic science.

Immunophenotyping of TCR and BCR clonotypes

Technology Development

POSTER 522

Alex Chenchik, **Michael Makhanov, Russell Darst, Tiangbin Liu, Lester Kobzik**

Cellecta, Inc., Mountain View, CA

TCR and BCR repertoire profiling holds great potential for understanding disease mechanisms and for development of new therapeutics in infectious diseases, autoimmunity and in immuno-oncology. However, this potential could be greatly improved by combining information about receptor clonotypes with immunophenotypes of T and B cells. To facilitate these studies, we developed a novel technology for combined profiling of all human TCR and BCR variable regions and phenotypic characterization of immune cells in bulk and at the single-cell level in PBMC and immune cell fraction samples.

The developed TCR/BCR Immunophenotyping method involves multiplex RT-PCR amplification and sequencing of CDR3 regions of TCR and BCR genes and a set of the most informative T- and B-cell phenotyping genes. Bioinformatics analysis of NGS data allows profiling of TCR/BCR clonotypes, identify major immune cell subtypes and their activation status.

Data will be presented showing how combined TCR/BCR clonotype analysis combined with targeted expression profiling of immune cells can be applied for large-scale discovery of novel cell typing and activation biomarkers in several immune-responsive model systems. Preliminary studies demonstrate the assay has unparalleled throughput, sensitivity, and improved cost-effectiveness for high-throughput immunity biomarker discovery applications.

Improving detection sensitivity for PacBio ISO-Seq using CRISPRclean™

Technology Development

POSTER 523

Jon Armstrong, **Josh Diaz, Azeem Siddique, Keith Brown**

Jumpcode Genomics, San Diego, CA

ISO-Seq is described as a long read sequencing method to detect and discover full length transcript isoforms. The raw base accuracy and read lengths of the Pacific Bioscience's Sequel II system enables a direct look at transcript isoforms without the need for inference through short read sequencing. The current capacity of a single Sequel II flow cell enables up to 4M reads, of which the top expressed genes take up the vast majority of read space, limiting the detection of low expressed transcript isoforms. In this presentation, we explore the use of CRISPRclean™ technology by applying it toward the depletion of abundant mitochondrial and ribosomal protein coding transcripts (polyA) from ISO-Seq libraries. CRISPRclean™ uses the in vitro CRISPR-CAS system to generate double stranded breaks in cDNA molecules during the library preparation step. The targeted molecules no longer have adapter sequences on both ends and are therefore reduced to background during library amplification. In this study, we designed and tested approximately 1,800 sgRNA molecules spaced every 50 bp across the full length sequences of 90 ribosomal and 10 mitochondrial derived mRNA sequences. The relative abundance of these gene targets went from approximately 40% of the sequencers unique read output to less than 1% of reads, shifting the remaining 39% toward more informative transcripts. In this poster we characterize the increased sensitivity to detect novel isoforms between depleted and non-depleted libraries from the same sample. As the CRISPRclean technology is applied post cDNA synthesis, the same library can be used to deplete additional transcripts with additional guides sets for further discovery. Results from a single cell version using Droplet based single cell barcoding and a modified ISO-Seq and CRISPRclean workflow are also presented.

Application of a new digital microfluidics platform to enable improved on-demand library preparation for long read sequencing

Technology Development

POSTER 524

Allyson Day[1], **Erin LaRoche**[1], **Katie Larkin**[1], **Tera Bowers**[1], **Niall Lennon**[1], **Stacey Gabriel**[1]

1 Broad Institute of MIT and Harvard, Cambridge, MA

Growing demand for long read sequencing technologies has increased the need for more library preparation approaches. The complexity and inefficiency of current manual workflows, in particular for whole genome sequencing applications, is cumbersome and could benefit from a straightforward streamlined approach. In response to these concerns, Broad's Genomic Platform is collaborating with Miroculus and Pacific Biosciences to develop a faster, more robust, and cost efficient automated library preparation. The Miroculus instrument, Miro Canvas, enables rapid hands free creation of HiFi libraries by reducing the potential for human error in manual processes. The microfluidic technology minimizes the shearing effects pipetting has on long DNA fragments. PacBio's latest chemistry improvement for HiFi reads, increases output data for detection of all variants with greater phasing and accuracy of single nucleotide polymorphisms and structural variants even in hard to reach areas of the genome leading to better data for clinically relevant human genomes.

As an initial test, we have demonstrated the feasibility of the technology. Sheared genomic DNA and reagents were loaded onto the Miro Canvas where the entire library construction process was done, including all enzymatic reactions, bead purifications, and incubations. The final output is a sequencer ready HiFi library for the Sequel IIe system. Minimal pipetting steps and straightforward setup resulted in less errors and more consistent library yields. We observed that this technology is as good as manual techniques in terms of turnaround time, reproducibility, and data quality. Although the current capacity is one sample, if used on a larger scale, the Miro Canvas has the ability to increase laboratory productivity while lowering hands-on processing time.

In the future, we aim to use this technology to lower sample input requirements and normalize library construction for all samples regardless of initial sample quality. This technology can also be applied to other long read sequencing applications such as PacBio's RNA sequencing application Iso-Seq to enable rapid, hands free creation of full length transcripts.

Long-read variant-robust full genome sequencing of SARS-CoV-2 using molecular inversion probes

Technology Development

POSTER 525

Eric D. Boyden[1], **Jack M. Amaral**[1], **Patrick C. Saunders**[1], **Gregory J. Porreca**[1]

1 Molecular Loop Biosciences, Inc., Alltson, MA

Effectively combating pandemics such as Covid-19 requires routine and systematic complete viral sequencing of positive samples to identify and track emergent variant strains. The widely used PCR-based ARTIC assay is effective but prone to mutation dropouts and failed amplification when new mutations are present at primer annealing sites, which has required multiple revisions of primer sequences. We previously presented an alternative variant-robust assay for short-read sequencing that uses heavily tiled molecular inversion probes (MIPs) to provide high levels of redundancy (average 7.5X tiling depth), which safeguards against amplification and mutation dropouts while also ensuring a high degree of coverage uniformity that facilitates high genome completion rates across a broad spectrum of Ct values.

Two limitations of short-read NGS assays are an inability to phase distant variants and a lack of sensitivity to identify larger indels, due to the difficulty of accurately aligning indel-containing reads to the genome. Long-read platforms by Pacific Biosciences and Oxford Nanopore Technologies can produce reads from cDNA libraries that are several hundred to thousands of bases in length, with far more power to phase variants and identify moderately sized indels (e.g. up to 0.5X read length). However, capitalizing on these key advantages requires libraries with sufficiently long template molecules, which is not a common feature among published MIP protocols. To establish that our technology is capable of producing long templates that are optimal for long read sequencing, we designed panels of MIPs that capture the SARS-CoV-2 genome in overlapping elements of either 675 or 1215 bp, with average tiling depths of 22.5X and 40.5X respectively, ensuring an extremely high level of probe redundancy. Additionally, our highly scalable and easily automatable assay requires less than 2 hours of hands-on labor in 4 addition-only steps including a combined reverse transcription + hybridization step, and may be performed using either a short hybridization in a single-day workflow for quick turnaround, or an overnight hybridization for slightly improved coverage.

This work represents a significant step forward in the expansion of MIP technology to new applications. We have shown that MIPs are capable of capturing much larger elements than they have historically been designed to target, and potentially even larger elements that will further maximize their utility in conjunction with long-read sequencing platforms.

Optimized enzymatic fragmentation workflow for the rapid construction of high-quality whole genome and whole exome libraries that is compatible with several blood collection methodologies

Technology Development

POSTER 526

Zane Jaafar[1], **Josh Haimes**[1], **Thomas D. Harrison**[1], **Lindsay Peterkin**[1], **Kristin Scott**[2], **Martin Ranik**[1], **Brian Kudlow**[1]

1 Watchmaker Genomics, Boulder, CO
2 Colorado State University, Department of Cell & Molecular Biology, Fort Collins, CO

Many state-of-the-art short read NGS library preparation workflows for whole human genome sequencing (WGS) utilize sonication of high molecular weight genomic DNA. Existing alternatives which utilize enzymatic shearing suffer from drawbacks including difficulty tuning insert sizes, sequencing artifacts and sensitivity to DNA input amount and quality. Crucially, the source tissue or body fluid and the DNA extraction methodology often have a significant impact on the effectiveness of enzymatic library preparation. We assessed the utility of a novel enzymatic fragmentation methodology for the preparation of WGS libraries from DNA extracted from a range of clinically relevant sample types, including dried blood spots and blood extracted using Tasso devices. Fragmentation and bead purification parameters were optimized to yield desired insert sizes from a variety of DNA sources and were compared to typical distributions generated by sonication. We titrated the input genomic DNA mass to establish the lower limit that enabled the reproducible construction of PCR-free libraries for sequencing on the NovaSeq 6000 system. Sequencing data quality parameters, such as coverage uniformity, minimal sequence context bias as well as presence of chimeric reads and other undesirable artifacts, were interrogated. Our data shows that this chemistry offers a highly tunable, single tube solution for the rapid preparation of PCR-free WGS libraries from human DNA extracted using various blood collection and purification strategies. Additionally, we tested the workflow in the context of whole exome sequencing (WES) and found that our solution enabled reproducible generation of high-quality WES libraries from a large range of DNA inputs (down to 10 nanograms) with minimal required library amplification cycles. Taken together, these data demonstrate the utility of this novel fragmentation chemistry for rapid generation of high-quality libraries from a variety of clinically-relevant samples.

Integrated nanoliter to microliter liquid handlers for streamlined and fully automated 384-well plate library preparations

Technology Development

POSTER 527

Soheila Beck, **Sudha Savant***, **Shilpa Parakh**, **Alex Pierson**, **Kristie Najir**, **Partha Banerjee**

Beckman Coulter Life Sciences, Indianapolis, IN
*Presenting author

The Illumina Nextera XT DNA Library Preparation Kit can prepare indexed paired-end libraries for sequencing on Illumina sequencing platforms. The kit uses only 1 ng of input DNA and fragments the DNA using an enzymatic reaction called tagmentation. The kit does not require mechanical shearing, which makes it a great fit for high-throughput NGS labs working with small genomes.

In this poster, we show time- and cost-saving data on the integrated Echo 525 and Biomek i5 Multichannel (MC) liquid handlers to automate and reduce the total reaction volume of the Nextera XT Prep Kit by up to 20 times, while maintaining data quality, and increasing the throughput by ~4 times. The miniaturized workflow requires as little as 50 pg of input sample. All miniaturized reactions are performed on an Echo 525 liquid handler; while source plate preparation, individual bead cleanups, thermocycling incubations, and shaking are handled by the Biomek i5 MC liquid handler with 384-well head, with an on-deck thermal cycler. The Biomek i5 MC liquid handler's 360˚ offset gripper accesses the Echo 525 liquid handler without the need for additional integration.

The entire automated workflow from aliquoting reagents to final, sequence-ready, mass-normalized and pooled libraries can be done in 4.5 hours for 384 samples on the Biomek i5 MC-Echo 525 integrated system. This is an at least 4 times faster processing time as compared to a workflow without an Echo liquid handler and a 384 head. The integrated system provides the ability to streamline and scale up for 384-well plate bead cleanups and increased walk-away time, as compared to an Echo 525 system alone. The 20-times miniaturization of the Nextera XT Prep Kit resulted in over 90% cost savings on reagents, and over 75% time savings compared to a method that only uses traditional transfer methods.

Unlocking the power of scale with high-throughput multiomic single-cell assays

Technology Development

POSTER 528

Dagmar Walter, **Qiang Gong, Luz Montescarlos, Ariel Royall, Jerald Sapida, Corey Nemec, Mohammad Rahimi**

10x Genomics, Pleasanton, CA

Understanding heterogeneity at the single cell level has enabled significant insights into biological mechanisms and their contributions in health and disease. But researchers may find themselves restricted by scalability challenges, limiting which questions they can answer.

Here, we use the Chromium X with the latest developments in high throughput (HT) single cell profiling, the Chromium Single Cell 3' Gene Expression HT and Chromium Single Cell Immune Profiling HT workflows. This allows us to scale our experiments 5-fold to nearly a million cells per run, allowing technical and cost barriers to be overcome, enabling ambitious experiments and empowering researchers to address the most complex biological questions.

Recent technological advancements have enabled multiomic approaches integrating cell surface protein, immune repertoire and antigen specificity measurements at single cell resolution, providing comprehensive, high-throughput characterization of immune cells and allowing a better understanding of the cell state by characterizing each layer. Using the 10x Genomics Single Cell Immune Profiling HT solution with Feature Barcode technology, we compared healthy and leukemic human bone marrow samples. The HT assay enabled us to easily scale up the number of cells profiled by 7-fold, going from approximately 8,000 cells to greater than 60,000 cells per sample. By drastically scaling up the number of cells profiled, we were able to identify rare cell populations, such as low-frequency hematopoietic stem cells, in the healthy bone marrow samples, and get more refined and accurate calling of rare leukemia subpopulations in the diseased sample. Additionally, the additional cell throughput allowed for the identification of rare (>3 cells) T cell receptor (TCR) clones that wouldn't have been identified with lower cell loads, providing novel insights into the immune response to the disease.

Increasing cellular throughput allows for increased biological resolution and critically, the identification of rare cell types.Further, it unlocks the potential to screen for rare TCR clonotypes and/or rare antigen-specific TCR sequences providing increased profiling power for the study of immune repertoire response to infection and disease. We envisage this technology providing key insights across various research areas, including MRD detection, deep immune repertoire analysis, and rare cell type discovery.

Plant protoplast scRNA-Seq (ppRNA-Seq) on the Nadia Instrument

Technology Development

POSTER 529

Alita Zabrecky[1]

1 Dolomite Bio, Royston, United Kingdom

Single-cell transcriptomic studies of leaf cells can provide insight into important aspects of plant biology, such as the photosynthetic pathways and gaseous exchange. However, large-scale transcriptomic studies of single protoplast cells from plant leaves, in comparison to animal models, are remarkably absent. If single-cell leaf transcriptomes are available, they are isolated through microdissection, which limits potential throughput.

Previously, the use of single-cell and microfluidic techniques on single protoplast cells was not feasible due to technological concerns such as cell size, the requirement of intense enzymatic digestion, and specific buffer conditions.

We assessed an optimised protocol on the Nadia Innovate for single-cell RNA sequencing of Arabidopsis thaliana leaf protoplasts. During protoplast encapsulation with high viscosity cell buffers on the Nadia, stable droplet formation and high throughput encapsulation of thousands of isolated plant leaf cells were successful. This was confirmed with real-time visual feedback. The integrated single-cell RNA-sequencing workflow for use with plant protoplasts (ppRNA-Seq) developed by Dolomite Bio generated high-quality transcriptomic data from a variety of cell types, in particular large mesophyll leaf cells, by microfluidic methods.

To conclude, the Nadia and Nadia Innovate platforms, alongside the new ppRNA-Seq protocol, enable efficient encapsulation and the generation of single-cell RNA-Seq datasets for hundreds of single plant leaf protoplasts. With these methodological advances in high-throughput microfluidic technology, pioneering opportunities on single-cell genomic workflows and the acceleration of plant leaf research in the single-cell field are possible.

Development of Exome 2.0, the best-in-class exome and improved target enrichment system

Technology Development

POSTER 530

Elaine Maggi[1], **Steve Oh**[1], **Michael Bocek**[1], **Leonardo Arbiza**[1], **Rebecca Nugent**[1], **Adriana Arneson**[1], **Derek Murphy**[1], **Esteban Toro**[1], **Siyuan Chen**[1]

1 Twist Bioscience, South San Francisco, CA

Target enrichment offers distinct advantages over whole genome sequencing. Specifically, the complexity and costs of sample analysis are reduced through focused sequencing of select regions of interest. Leveraging target enrichment, whole exome sequencing has become a powerful and well established tool used to interrogate the coding regions of the genome. There are several target enrichment systems and exome capture panels available on the market, including options from Twist Bioscience. Here we present data from our most up-to-date whole exome capture solution. Featuring an improved hybridization system and optimized capture panel design, Exome 2.0 enables best-in-class performance with options for same day or overnight target enrichment workflows.

Our newly designed panel has been expanded to include an exhaustive list of regions of clinical significance. This additional content includes ClinVar coverage of over 99% representing an appreciable improvement over previous exomes. The probe design has also been rebalanced and specifically tailored to produce increased uniformity of coverage on the Illumina Novaseq sequencing platform. In addition to these improvements, the updated overnight hybridization system reduces undesirable off-bait by over 50% compared to previous offerings. This increased uniformity of coverage combined with an improved on-target rate results in lower overall sequencing needs and costs per sample.

Further improvements include the addition of a bead-based alternative to the speedvac drydown step for both the overnight and single-day hybridization workflows. The configuration of the hybrid-capture system also allows for customization of targeted regions and the ability to multiplex up to 16 samples in a single capture reaction.

We demonstrate that Exome 2.0 has improved performance in several key areas including uniformity of coverage and percent on bait. These results combined with our flexible workflow makes our system an ideal option for whole exome sequencing.

Novel approach toward single-cell RNA-seq of PFA-fixed cell samples

Technology Development

POSTER 531

Jing Zhou, **Janaki Acharya**, **Zhengyue Li**

Singleron Biotechnologies, Inc., Woodbridge, CT

Current high throughput scRNA-seq technologies usually require fresh cells with high viability, which means the collected samples have to been processed immediately once they are obtained. Although single nucleus RNA sequencing has been demonstrated as an alternative for frozen samples, these methods are yet to be fully compatible with clinical archival cell samples, for example, fixed with paraformaldehyde (PFA) or formalin. Thus, current scRNA-seq methods are fundamentally limited due to challenges in sample collection, transportation, and storage, which may also increase the cost associated with sample pre-processing for scRNA-seq. Therefore, it would be highly desirable to develop new methods that carefully compatible with archival cytologic specimens, such as blood or bone marrow aspirates that have been treated with a fixative such as formaldehyde. Herein we report on a proprietary technology platform named Multiplexing by Cleavage, Delivery, and Barcoding (MX-CDB) developed in Singleron Biotechnologies for high-throughput single-cell RNA sequencing of PFA-fixed cell samples. We successfully extracted high-quality mRNAs from these samples that were pre-fixed with 4% PFA, generated barcoded single-cell transcriptome library, and performed paired-end sequencing to obtain single-cell transcriptome data with the quality comparable to that obtained from fresh cell samples. We investigated the impact of long term storage of PFA-fixed cells by comparing the data quality from samples stored at -20 oC for one week, two weeks, and found there is no significant degradation of RNAs in the cell samples using this new approach. Singleron's MX-CDB platform provides a simplified and widely applicable workflow that greatly improves the compatibility of scRNA-seq with clinical cytologic samples by demonstrated capture and barcoding mRNA from thousands of single cells in fully fixed biological samples to perform high-throughput single-cell sequencing.

Utilizing an alternate cell fixation method to expand opportunities for single-cell RNA-sequencing

Technology Development

POSTER 532

David Corney, **Yang Han, Yu Qiu, Cong Shen, Yongjun Fan, Michael Stephens, Christopher Mozdzierz, Haythem Latif, Ginger Zhou**

GENEWIZ, South Plainfield, NJ

Traditional approaches to performing transcriptional profiling have relied heavily on bulk RNA-seq which, while straight-forward and cost-effective, eliminates all evidence of heterogeneity that may have been present within a complex and diverse tissue. For this reason, high-throughput droplet- and microfluidic-based approaches for performing single-cell RNA-seq (scRNA-seq) have become powerful techniques to answer important questions in a wide range of fields. However, a significant bottleneck to widespread and routine usage has been the specialized and involved sample preparation requirements prior to performing scRNA-seq. Typically, tissues must be dissociated and processed immediately, which makes it difficult to batch and transport samples off-site, or cryopreserved, which is associated with challenges related to poor cellular viability upon sample thawing.

To circumvent these challenges, we and others have reported alternative approaches such as cryopreservation of tissues prior to dissociation, methanol-based fixation of cells, and performing single nuclei RNA-seq. To further broaden the ability for researchers to deploy scRNA-seq, we report an alternative sample preparation and processing workflow which offers benefits compared to currently utilized approaches. Here, we describe our experience deploying a pre-commercial workflow developed by 10x Genomics to resolve these sample management challenges. On-site paraformaldehyde (PFA) fixation allows samples to be collected and transported to off-site laboratories without sacrificing integrity or data quality, which creates new possibilities for sample accessibility, throughput and batched analysis. This approach also allows for analysis of transcriptome-wide gene expression at high sensitivity, with simultaneous cellular readouts. We will report case studies comparing this new approach to standard approaches and highlight the advantages and possibilities that it enables.

Tailed multiplex PCR development permitting SARS-CoV-2 surveillance sequencing at scale in the United Kingdom

Technology Development

POSTER 533

Naomi Park[1], **Emma Betteridge**[1], **Scott Goodwin**[1], **Lesley Shirley**[1], **Keith James**[1], **Ian Johnston**[1]

1 The Wellcome Sanger Institute, Hinxton, Cambridge, United Kingdom

In March 2020 the Wellcome Sanger Institute (WSI) pivoted it's world leading expertise in sample logistics, sequencing pipelines, metadata handling, bioinformatic analysis and interpretation to the large-scale sequencing of SARS-CoV2 extracts sent from national diagnostic testing centres.

To prepare the samples for Novaseq sequencing, Artic V3 multiplex PCR designed by the Artic network was dovetailed with existing adapter ligation library construction methods in use at WSI. Within weeks a pipeline capable of generating several hundreds of high quality genomes per day was established. However, the process was limited in its scalability to the throughput required as the pandemic progressed. Restrictions on space, FTE, large liquid handler availability combined with complexity of the process and competing WSI demands meant the scale of sample volumes required to inform public health response could not be met.

The rapid development of a simplified tailed Artic V3 based multiplex library construction process substantially reduced turnaround time and costs without compromise of data quality. In the midst of a global shortage of plasticware the refinements and improvements also markedly reduced consumable requirements. This process enabled a step change for the delivery of SARS-CoV-2 genomes, effectively doubling throughput overnight.

Here we describe:

- Challenges during tailed multiplex PCR development
- The use of Oxford Nanopore sequencing for rapid feedback on effectiveness of optimisations
- Additional process improvements including primer rebalancing for even genome coverage

Tailed multiplex PCR SARS-CoV2 library construction for IIllumina has been in use at WSI since November 2020.
To date,~ 750,000 genomes have been sequenced using the tailed library construction process.
COVID-19 ARTIC v3 Illumina library construction and sequencing protocol - tailed method (protocols.io)
https://www.sanger.ac.uk/science/covid-19-science/
https://covid19.sanger.ac.uk/lineages/raw

Technology Development

POSTER 534

Extraction and library preparation of ultra-long high-molecular-weight gDNA for nanopore sequencing

Luo Sun[1], **P.A. Giron Koetsier**[1], **Danielle Freedman**[1], **Eileen T. Dimalanta**[1], **Eric J. Cantor**[1], **Theodore B. Davis**[1]

1 New England Biolabs, Ipswich MA

Extraction of pure and ultra high molecular weight (UHMW) gDNA is crucial for Ultra-long (UL) read Nanopore sequencing. Currently there are several methods that can extract high quality HMW gDNA but these methods are either time-consuming or costly. Here, we demonstrate a novel glass bead-based extraction method using the Monarch HMW DNA Extraction Kit that generates high quality, mega-base long gDNA and a bead cleanup protocol during library prep from a human cultured cell line and blood. Nanopore sequencing results indicated a sequencing length of N50 >100kb and >90kb, respectively, with the longest read >2 Mb. Also, by simply changing the agitation speed during incubation of beads and cell lysis, we can generate a desired range of 50kb up to mega-base long DNA fragments. This provides a simple, low-cost method that avoids the requirement for DNA fragmentation and enables flexibility in DNA length.

In addition, we set out to develop a short protocol for alcohol-free cleanup of UHWM DNA libraries based on glass-bead binding. Employing a binding buffer containing hexamine cobalt chloride, we have established a 30-minute convenient and high-yield cleanup protocol. With 30-90 minutes homogenization time supported by occasional gentle pipetting with a wide bore pipette tip, homogeneous UL DNA libraries are obtained that are ready to be loaded on the flow cell.

In summary, we are demonstrating results on a new workflow with a same day preparation time, resulting in high DNA yield, purity, and high-quality UL sequencing results from samples isolated from cultured human cells and human blood.

Technology Development

POSTER 536

Robust detection of translocations in lymphoma FFPE samples using targeted locus capture-based sequencing

Amin Allahyar[1], **Mark Pieterse**[1*], **Joost Swennenhuis**[2*], **G. Tjitske Los-de Vries**[3*], **Mehmet Yilmaz**[2], **Roos Leguit**[4], **Ruud W.J. Meijers**[4], **Robert van der Geize**[5], **Joost Vermaat**[6], **Arjen Cleven**[7], **Tom van Wezel**[7], **Arjan Diepstra**[8], **Léon C. van Kempen**[8], **Nathalie J. Hijmering**[3], **Phylicia Stathi**[3], **Milan Sharma1**, **Adrien S.J. Melquiond**[1], **Paula J.P. de Vree**[1], **Marjon J.A.M. Verstegen**[1], **Peter H.L. Krijger1**, **Karima Hajo**[2], **Marieke Simonis**[2], **Agata Rakszewska**[2], **Max van Min**[2], **Daphne de Jong**[3], **Bauke Ylstra**[3], **Harma Feitsma**[2], **Erik Splinter**[2#], **Wouter de Laat**[1#]

1 Oncode Institute and Hubrecht Institute, KNAW and University Medical Center Utrecht, Utrecht, The Netherlands
2 Cergentis BV, Utrecht, the Netherlands
3 Amsterdam UMC-Vrije Universiteit Amsterdam, Department of Pathology and Cancer Center Amsterdam, Amsterdam, The Netherlands
4 University Medical Centre Utrecht, Department of Pathology, Utrecht, The Netherlands
5 Laboratorium Pathologie Oost-Nederland, Hengelo, The Netherlands
6 Leiden University Medical Center, Department of Hematology, Leiden, The Netherlands
7 Leiden University Medical Center, Department of Pathology, Leiden, The Netherlands
8 University of Groningen, University Medical Centre Groningen, Department of Pathology & Medical Biology, Groningen, The Netherlands
*Contributed equally
#Communicating Authors

In routine diagnostic pathology, cancer biopsies are preserved by Formalin-Fixed, Paraffin-Embedding (FFPE) procedures to be examined for (intra-) cellular morphology. Such procedures inadvertently induce DNA fragmentation, which compromises sequencing-based analyses of chromosomal rearrangements. Yet, rearrangements drive many types of hematolymphoid malignancies and solid tumors, and their manifestation is instructive for diagnosis, prognosis and treatment. Here, we present FFPE-Targeted Locus Capture (FFPE-TLC) for targeted sequencing of proximity-ligation products formed in FFPE tissue blocks, and PLIER, a computational framework that allows automated identification and characterization of rearrangements involving selected, clinically relevant, loci. FFPE-TLC, blindly applied to 149 lymphoma and control FFPE samples, identifies the known and previously uncharacterized rearrangement partners. It outperforms Fluorescence in situ hybridization (FISH) in sensitivity and specificity and shows clear advantages over standard capture-NGS methods, finding rearrangements involving repetitive sequences which they typically miss. FFPE-TLC is therefore a powerful tool for accurate targeted rearrangement detection in FFPE specimens.

SARS-CoV-2 cDNA library construction using a reduced reaction volume protocol on the Mosquito HV Genomics and Dragonfly Discovery nanoliter liquid dispensing instruments

Technology Development

POSTER 537

Yanping Zhang[1], **Brandon Pate**[1], **Xiaohui Zhou**[1], **David Moraga**[1], **Alberto Riva**[1], **Samantha Lee**[2], **Steven J. Madore**[1]

1 Interdisciplinary Center for Biotechnology Research, University of Florida, Gainesville, FL
2 SPT Labtech Ltd., Melbourne, United Kingdom

The identification of SARS-CoV-2 viral variants remains a critical step in controlling the future spread of infection. Using genome sequencing from isolated viral RNA, it is possible to detect and quantify circulating viral lineages (including novel, potentially more transmissible and/or pathogenic) in the human population. This has important implications regarding the surveillance of potentially more infectious and dangerous viral variants and the ability to proactively implement specific public health measures to help control variant spread. To meet the demand for high throughput genomic sequencing, the ICBR miniaturized a standard protocol for SARS-CoV-2 genome library construction using the SPTLabtech Mosquito® HV Genomics and Dragonfly® Discovery liquid handling platforms with the Illumina COVIDSeq emergency use authorized Test protocol.

We first manually tested reducing the library synthesis reaction volume to half and quarter volumes with different inputs amounts of total RNA. Results showed that in comparison to the full volume reaction, DNA library product of sufficient quality and quantity can be consistently generated with 10ng input RNA at ¼ reduced reaction volume (RRV). Adapting the COVIDSeq library prep onto the SPTLabtech Mosquito® nanoliter liquid dispensing instrument allowed further reaction volume reduction to 1/5 scale.

To assess what effects if any the RRV protocol has on DNA sequencing, 384 RNA sample obtained from SARS-CoV-2 positive patients were split into two separate groups. One group was processed manually using the full-scale reaction protocol and the second group was processed on the Mosquito using the 1/5 reaction volume. DNA sequencing was performed on the Illumina NovaSeq6000 on the S1 flow cell using the 2 x 100 configuration. The sequencing results showed that no difference in SARS-CoV-2 genome coverage or in variant identification rate. The RRV protocol using the Mosquito® HV Genomics and Dragonfly® Discovery facilitates the generation of 384 high quality COVIDSeq libraries per day at a significant reagent cost savings.

Increasing the throughput of NGS sequencing through rational design and empirical validation of >3,000 Unique Dual Indexes (UDIs)

Technology Development

POSTER 538

Richard Gantt[1], **Adriana Arneson**[1], **Lydia Bonar**[1], **Palak Sheth**[1], **Esteban Toro**[1]

1 Twist Bioscience, South San Francisco, CA

NGS libraries are typically prepared by isolating nucleic acids to a desired length and ligating specialized adapters to both fragment ends. Barcodes are included in the adapters to serve as a unique identifier so many individual samples can be sequenced together. The barcodes are added either as part of the adapters for ligation, or via amplification with barcoded primers after ligating universal adapters to each sample. The universal adapter approach is more scalable, yet the number of available UDI primer pairs remains limited due to constraints in barcode design and validation. In an era of ever-increasing sequencing capacity, this unnecessarily limits sample throughput and increases the cost per sample.

To address these limitations, improvements to both barcode design and screening validation were developed. The computational design of barcode sequences grows in complexity along with the total number of unique barcodes intended for use in a single sequencing run. Barcode sequences must be selected while keeping two distinct criteria in mind. First, the frequency of DNA base utilization must remain balanced to ensure barcode choice does not introduce bias during sequencing. Second, barcodes must be sufficiently different from one another to minimize the risk of confusing one barcode for another during sequencing. Barcode design can be difficult to scale, primarily because with each new barcode candidate, the algorithm must measure the distance to all the other current barcode candidates. Additionally, each barcode must be assessed for risk of off-target interaction both with itself and potential hybridization with other barcodes. Our design approach distributes all of these calculations across a cluster of computing resources allowing for the exploration of possible good candidate barcodes and a ranking of qualifying barcodes by both their distance metric and their risk of off-target hybridization. The approach yielded more than 10,000 pairs of 12bp UDI primer candidates for commercialization.

To assess the performance of individual barcodes we highlight the development of Artificial Barcoded Libraries (ABLs), a high throughput NGS library generation method that utilizes fragmentation of a human genomic source and automation to generate thousands of unique NGS libraries with minimal bias. Performance criteria for a uniform UDI primer set with >3,000 members that are suitable for high multiplex hybridization applications and investigations are presented.

Real-time single-molecule sequencing of nucleosomal DNA and combinatory histone modification detection with zero-mode waveguides

Technology Development

POSTER 539

Pengyu Zheng[1,2], **Meni Wanunu**[1], **Efrat Shema**[2]

1 Department of Physics, Northeastern University, Boston, MA
2 Department of Biological Regulation, Weizmann Institute of Science, Rehovot, Israel

Nucleosome, the building block of chromatin, contains many factors and information apart from the 147bp DNA that wraps around it, such as the modification on the histone tail which has been proven to play an important role in nucleosome dynamics and gene regulation. The prevalent method for measuring the prevalence of a certain histone modification across the genome is Chromatin immunoprecipitation sequencing (ChIP-seq), which has limitations in sensitivity, throughput, and signal-to-noise ratio as well as the inability to identify individual nucleosomes with combinatorial modifications such as the bivalent nucleosomes. We here present high-throughput single molecule imaging and detection of multiple histone modifications on mono-nucleosome with zero-mode waveguide (ZMW) and subsequent sequencing of the nucleosomal DNA to locate the genomic position of individual nucleosome with the modifications. This serves as proof of principle that single molecule real-time (SMRT) sequencing, a third-generation DNA sequencing platform, can work in combination with both histone modification detection and DNA sequencing to decipher multiple layers of nucleosomal epigenetic codes.

Access to synthetic oligos produced by Enzymatic DNA Synthesis (EDS) accelerated the development of Candida glabrata mutants during the COVID-19 pandemic

Technology Development

POSTER 540

Colin Clairet[1*], **Steven Quistad**[2], **Xavier Godron**[2], **Christophe D'Enfert**[1]

1 Fungal Biology and Pathogenicity, Institut Pasteur, Paris, France
2 DNA Script, Le Kremlin-Bicêtre, France
*Presenting author

Candidemia occurs when members of the fungal genus Candida infect a patient's bloodstream. Treatment of candidemia is particularly difficult in immunocompromised patients for several reasons, including biofilm formation and resistance to antifungal compounds. Two major classes of antifungals that are currently used for the treatment of invasive fungal infection (IFI) include echinocandins and azoles. Candida glabrata, the second most prevalent species responsible for IFI, is able to rapidly acquire resistance to echinocandins and is naturally tolerant to azoles. The development of novel antifungal compounds is therefore urgently needed.

Protein kinases are central to many metabolic processes, including chemical sensing and adaptation; however fewer than 30% of the predicted kinases have been functionally characterized in fungal pathogens. In this study we used molecular approaches to identify and characterize protein kinases involved with antifungal resistance. Kinase-encoding genes were systematically deleted and the sensitivity of the resulting mutants to existing antifungal compounds was determined. Like most reverse genetic approaches, the protocol required access to synthetic oligos.

When this study was launched, standard phosphoramidite oligonucleotides for the generation of Candida mutants were ordered from a commercial vendor. After the onset of the COVID-19 pandemic, access to synthetic oligos was severely impacted, with delays of up to three months for standard PCR primers. The SYNTAX System (DNA Script) was made available to us in February 2021, providing access to oligonucleotides produced with DNA Script's novel Enzymatic DNA Synthesis (EDS) technology. With the SYNTAX System, we were able to obtain next-day PCR primers that performed equivalently to the phosphoramidite primers used at the beginning of the study. This eliminated a major experimental bottleneck and accelerated experimentation. In addition, access to SYNTAX enabled us to generate Candida mutants for the remainder of the study. The SYNTAX System and Enzymatic DNA Synthesis (EDS) provide an alternative to standard oligonucleotide production that allows on premise, highly adaptable primer synthesis and has the potential to accelerate research programs that require access to oligonucleotides.

Keywords: Enzymatic DNA Synthesis (EDS), phosphoramidite chemistry, SYNTAX, Candida, antifungal resistance

Simultaneous RNA and protein imaging with a single-cell spatial omics platform reveals cellular composition changes in neurodegeneration mouse models

Technology Development

POSTER 541

Marcos Otero-Garcia, Natalya Kotova*, Irene Oh, Tarif Awad

Rebus Biosystems Inc., Santa Clara, CA
*Presenting author

Cell-type composition and expression of immune response genes are two key variables in neurodegeneration mouse model studies. Single-cell RNA sequencing provides a powerful approach to study these variables, at the expense however of critical anatomical and morphological information. Here we present a methodology which allows accurate gene expression quantification without compromising spatial resolution. Using the Rebus Esper™ automated high-throughput spatial omics platform, we quantified the expression of 30 target genes at the single-cell level, in addition to two proteins of interest, in the hippocampus of two disease models plus a wild-type control mouse. We were able to define the major neuronal and glial cell types, anatomical structures, and quantify expression changes in disease-related genes. We measured changes in the relative abundances of the defined cell types and their spatial relationships between disease conditions. We also explored the relationship between these pathological gene-expression changes and the presence of disease-related proteins. These results demonstrate the ability of the Rebus Esper spatial omics platform to yield high-throughput spatial omics data with single cell resolution. The combination of large imaging area, high-plex RNA quantification, and protein imaging offers a unique and powerful tool to explore complex tissues with an unprecedented level of detail.

An evaluation of the Ultima Genomics sequencing platform: scalable, high-throughput sequencing for low-cost whole genome sequencing

Technology Development

POSTER 542

Matthew Coole[1], Tom Howd[1], Florian Oberstrass[2], Martin Sosa[2], Sophie Low[1], Tyson Clark[2], Andrew Bernier[1], Bharathi Gandi[1], Jack Stohlman[1], Megan Shand[1], Corey Nolet[1], Yossi Farjoun[3], Niall Lennon[1], Stacey Gabriel[1]

1 Genomics Platform, Broad Institute of MIT and Harvard, Cambridge, MA
2 Ultima Genomics, Newark, CA
3 Lady Davis Institute for Medical Research, Montreal, Quebec City, Canada

The cost for Human Whole Genome Sequencing (WGS) has stabilized at a little less than $1,000 over the past few years, but the path to significant reduction is not clear. Lower WGS costs offer the chance to dramatically increase the production of population-scale studies and ultimately drive discovery and demonstration of utility of such data in the healthcare system. Ultima Genomics (UG) is a platform designed to enable continual cost decreases from a starting point of $1/Gb and operate at an industrial scale. Here we introduce and evaluate the performance of the UG novel DNA sequencer in our lab at the Broad Institute and demonstrate an automated, production-ready library construction workflow.

In order to operationalize the platform, we leveraged an off the shelf library preparation kit, coupled with Ultima Genomics indexed adapters, to generate PCR-Free libraries. After library quantification, normalization, and pooling, samples undergo a proprietary process to anneal library pools to sequencing beads. We evolved a highly manual sample preparation process into a robust and automated workflow which results in a final bead-bound sequencing library.

Sequencing was performed on UG sequencers, which utilize mostly-natural bases in a single-read, sequencing by synthesis process. To achieve a target read length near 300bp, sequencing runs were 444 flows (111 cycles across the four nucleotides). Our early version of the UG sequencer has an overall runtime (including instrument setup) of 24 hours and consistently generates enough data to drive 12 samples to > 30x mean genome coverage (average = 40.1x) per run. In order to benchmark performance, we sequenced 200 samples from the HapMap project to 30x mean coverage. Utilizing a new sequencing analysis pipeline developed to analyze the single-read, flow-based nature of the UG sequencing systems, we observe high quality SNP and short INDEL calling over the High Complexity Regions of the genome, with lower quality for long homopolymers and in Low Complexity Regions. In conclusion, we have developed an end-to-end workflow for generating PCR Free WGS data on the Ultima platform in a scalable manner that is suitable for large scale population studies and demonstrated results with early versions of chemistry and instrumentation with a path to further improvements in cost, quality, and throughput.

An evaluation of enzymatic fragmentation in the library preparation workflow: the Watchmaker Genomics reagent kit

Technology Development

POSTER 543

Marissa Gildea[1], **Sam DeLuca**[1], **Tom Howd**[1], **Zane Jaafar**[2], **Martin Ranik**[2], **Niall Lennon**[1], **Stacey Gabriel**[1]

1 Genomics Platform, Broad Institute of MIT and Harvard, Cambridge, MA
2 Watchmaker Genomics, Boulder, CO

As the demand for exome and whole human genome sequencing (WGS) continues to grow, evaluations of both workflow changes and novel technologies will lead advances in scientific development while ensuring competitive market pricing. Specifically, there is a need for the sample preparation process to evolve in order to match the dropping cost of sequencing. DNA fragmentation is a fundamental step in the library construction (LC) process, and mechanical shearing has historically been the frontrunner in workflow choice given its naturally unbiased and uniform performance. Recently, however, advances in enzymatic fragmentation workflows have been introduced as a truly competitive alternative to the mechanical shearing-based library preparation workflows. Here we present data on our evaluation of the Watchmaker Genomics library preparation reagent kit. Being a new workflow and library preparation method, our objectives were to: 1) Establish baseline relationships between fragmentation time, SPRI concentration, and DNA insert size; 2) Evaluate GC bias plots to understand the advancement in non-biased enzymatic fragmentation; 3) Compare standard Picard sequencing metrics (chimeric rates, coverage uniformity, etc.) to current WES and WGS workflows used in the Broad Institute's Genomics Platform; 4) Perform variant calling analysis on PCR Free libraries at 30x mean coverage and assess SNP and InDel specificity and sensitivity.

Watchmaker Genomics offers an extremely easy to use kit, capable of producing high quality libraries in under 2 hours. With a user only required to create a two component Frag/AT master mix and add indexed adapter to a ready-to-use ligation master mix, this streamlined workflow has the potential for a more efficient and cost-effective process. Through an iterative approach, we were able to optimize library preparation conditions to achieve the same high-quality sequencing data that were already established with mechanical shearing. Notably, the team was able to reduce chimeric rates below 0.5%, while maintaining mean insert size at 450 bp. We will present in detail a full comparison of sequencing results between libraries generated with the Watchmaker Genomics PCR Free kit and libraries generated with the clinically validated PCR Free WGS process utilized by the Broad Institute Genomics Platform.

Overcoming limitations of RNA-Seq library construction from FFPE samples using a novel workflow

Technology Development

POSTER 544

Elizabeth Dreskin[1], **Linda Lingelbach**[1], **Nishwetha Kumar**[1], **Cissy Jiang**[1], **Angelica Olcott**[1], **Wei Vivian Fu**[1], **Deanna Woo**[1]

1 Bio-Rad Laboratories, Hercules, CA

Formalin-fixed, paraffin-embedded (FFPE) tumor tissue is a common biospecimen for oncology research and is proposed as a source material for RNA sequencing to evaluate gene expression signatures in tumors. However, RNA purified from FFPE samples is typically highly degraded and of poor quality and therefore suboptimal as input material for RNA-Seq library construction. To overcome this limitation, we evaluated a novel workflow that utilizes a proprietary enzyme that bypasses the bias and limitations of reverse transcriptases (RT) and ligases, coupled with a post-library ribodepletion strategy that affords the benefits of retaining more of the transcriptome. The goal was to provide guidance on analytical methods to obtain reliable and accurate detection of biologically relevant gene expression using RNA-Seq from FFPE samples to perform transcriptome analysis.

This study examined FFPE RNA extraction methods from 4 breast cancer samples versus a matching fresh frozen (FF) sample using 4 commercially available kits to evaluate quality metrics such as Agilent RNA integrity number (RIN), Illumina DV200 measurements and the recovery of small RNAs during NGS library construction.

Our results demonstrate that this novel workflow generates a richer data set that represents more of the transcriptome, including both small RNAs (<200 base pairs) and long RNAs from FFPE samples. Gene discovery from FFPE samples exhibited greater concordance with matched fresh frozen samples when libraries were constructed using this novel workflow compared to libraries constructed with standard RT-ligase workflows. In addition, the strategy of pooling multiple barcoded libraries into one depletion reaction lowers the overall cost of library construction and normalizes data output, while maintaining the depletion efficiency the same as individual depletion. Taken together, the excellent analytical performance of this novel library preparation workflow supports its use as a reliable RNA-Seq expression method for common archival FFPE specimens in oncology research.

HIVE scRNAseq: the complete solution from single cells to libraries for RNA-seq, with integrated sample storage

Technology Development

POSTER 545

Elizabeth Sergison[1], **Todd Gierahn**[1], **Jim Flanigon**[1]

1 Honeycomb Biotechnologies, Weston, MA

Biological resolution at the level of individual cells is powering the next phase of precision health. Honeycomb Biotechnologies offers a new solution for single-cell technology: HIVE scRNAseq. The HIVE scRNAseq Solution integrates sample storage and single-cell profiling into a single workflow. The HIVE is a portable, handheld, single-use device that enables gentle capture, robust storage, and easy processing for the analysis of single-cell samples. The design of the HIVE is ideal for the recovery of fragile cells, loading of large sample volumes, and efficient processing of numerous samples in parallel.

The HIVE has an array of 65,000 picowells, with barcoded 3' transcript capture beads, for gentle loading and isolation of single cells. Once cells are loaded into the HIVE, a cell preservation solution is added, locking in the molecular signals. Cell-loaded HIVEs can now be stored or shipped until ready for Honeycomb's simplified and scalable HIVE processing and library preparation workflow. When reading of the biologic information is desired, reversible sealing of the HIVE's semi-permeable membrane is used to enable diffusion of a strong lysis solution followed by an optimized hybridization solution into the wells, leading to robust solubilization of cells while retaining all biologic macromolecules in each well for capture during the hybridization step. Beads with captured transcripts are then extracted from the HIVE by centrifugation, and the remaining HIVE library preparation steps are conducted in a 96-well plate format. Libraries can be sequenced, and primary analysis completed with Honeycomb's software solution specifically designed for the HIVE scRNAseq libraries.

Honeycomb aims to expand single-cell opportunities across basic, translational, and clinical research throughout the world. HIVE scRNAseq is now commercially available through Honeycomb Biotechnologies or PerkinElmer.

Reducing technical dropout events in single-cell RNA-seq through experimental depletion

Technology Development

POSTER 546

Jon Bezny, **Azeem Siddique, Keith Brown, Jon Armstrong**

Jumpcode Genomics, San Diego, CA

Regardless of method, single cell RNA-seq only captures a small fraction of the transcriptome of each cell. Often, this is due to inherent limitations of the methodology as reads 'dropout' at each step of library preparation. These dropout events are then confounded with noise, outliers, and stochastic genetic variation, resulting in the daunting computational task to parse out the true signal. Almost all computational algorithms have evolved to address this zero-inflation issue through a multitude of approaches, typically through various dimensionality reduction or imputation techniques. While consensus for a standardized computational approach has yet to be met, we present a turnkey experimentalmolecular solution that drastically reduces dropout events attributable to technical noise, statistically enhancing biological interpretation.

Traditionally, single cell data processing incorporates certain filtering and normalization steps prior to canonical clustering and downstream interpretation. Instead of removing those reads in-silico, our universally incorporated molecularexperimental solution removes those reads in-vitro, redistributing sequencing clusters to unique biologically relevant transcripts. Our method, called CRISPRclean, seamlessly incorporatesis seamlessly incorporated into any single cell RNA-seq workflow prior to the final PCR amplification. CRISPRclean, and leverages CRISPR Cas9 depletion and a custom designed guide library to remove reads that were previously never incorporated in downstream analysis. For example, by analyzing a cohort of publicly available single cell 10x data from various sources, roughly 30-50% of reads aligned to the genome but not the transcriptome, and thus, are conventionally ignored. By tailoring guides to deplete these genomic intervals in addition to the highest expressed protein coding ribosomal and mitochondrial genes, we have exhibited the ability to redistribute 50% of reads through in-silico depletion across samples representing 14 tissue typessingle cell data from 14 tissue types.

As an initial proof of concept, we performed our preliminary single cell depletion on a cohort of patient derived endothelial cells from coronary and pulmonary arterial beds. As a result of depletion, 30% of reads were re-distributed, resulting in a net 26% gain in additional unique molecular identifiers (UMIs) on a per cell basis. By leveraging existing algorithms such as molecular cross validation, affinity-based dispersion methods, and random matrix theory, we demonstrate the power of depletion to reduce technical dropouts, enhance the number of genes contributing to biological signal by 858 genes on average, while simultaneously mitigating any reduction in variation. The added benefit of performing depletion translated into the discovery of a biologically relevant phenotype illuminating the transcriptional disposition of coronary arterial beds toward atherosclerosis.

Rebus Esper: an integrated spatial omics platform with advanced optics, automated fluidics, and flexible assay chemistry

Technology Development

POSTER 547

Gabriel Mercado, Natalya Kotova, Marcos Otero-Garcia, Tarif Awad

Rebus Biosystems Inc., Santa Clara, CA

Traditional cell biology and microscopy recently merged with genomics to create spatial omics, a powerful new approach for studying biology. A range of methods and techniques are emerging, most of which are targeted to one class of biomolecules, often compromising sensitivity, specificity or dynamic range in favor of higher plex. Rebus Biosystems has developed the Rebus Esper™, an integrated platform that combines resolution-enhancing Synthetic Aperture Optics (SAO) with automated fluidics to enable multiple types of assays for detecting biomarkers at single-molecule, subcellular resolution. The streamlined workflow minimizes hands-on time and reduces variability by automating the laborious steps required for multiplexed single-molecule detection. Several assay chemistries will be available for the Rebus Esper platform. The Esper High Fidelity Assay for spatial transcriptomics automates single-molecule FISH to quantify 30 custom genes across ~1 cm^2 of tissue in ~2.5 days. Upcoming assays will combine dual detection of protein staining with quantification of RNA expression, increase plex levels to hundreds of genes, and enable the use of FFPE tissue, a challenging but important sample type for clinical and translational research. A flexible and open data analysis pipeline is integral to the Rebus Esper platform. Data can be processed and viewed using the provided tools, which can be integrated with publicly available or commercial analysis packages. With advanced optics, automated fluidics, and intuitive software, the Rebus Esper makes spatial omics accessible to any biologists seeking to visualize the molecular complexity of life.

Speed with precision: faster and longer cDNA synthesis accompanied by rapid high-fidelity PCR enabling long-range sequencing of RNA viruses

Technology Development

POSTER 548

Subrata Panja, Ryan C. Heller, Brian Komorous, David Schuster

Quantabio, Beverly, MA

RNA viruses are pathogens of critical concern and responsible for recent pandemics including H1N1 influenza, Ebola, Zika and COVID-19. RNA virus replication is an error-prone process often resulting in mutations that alter clinical features such as transmissibility and virulence. Rapid and accurate sequencing of RNA virus genomes remains a critical need for characterization of novel virus infections, strain surveillance and variant detection.

A first step in many RNA sequencing methods is conversion of the RNA genome into cDNA using a reverse transcriptase (RT). Despite its pivotal role, many commercially available RTs have a limited ability to generate long cDNA products because of low processivity and poor activity at temperatures sufficient to relax complex secondary structures often adopted by RNA viruses. Short-read NGS methods often rely on cDNA synthesis using random primers followed by high-order multiplex PCR that generates amplicons for sequencing. Development of validated multiplex PCR panels for reliable pathogen surveillance is challenging and laborious. This methodology often struggles with comprehensive mutation calls in a rapidly mutated environment or fails if the pathogen adopts larger structural variations including deletions, duplications, or insertions.

At Quantabio, through a program of directed mutagenesis and enzyme variant screening, we engineered a novel RT called qScript Ultra with enhanced thermostability, velocity, and processivity. In this study we used RNA from the large OC43 coronavirus as a model system and examined the ability of this novel RT to synthesize very long cDNA of up to 20-30 kb for template generation in long-read RNA sequencing technologies. Using a single gene-specific primer at the 3'-end of the OC43 RNA we were able to synthesize cDNA spanning the entire 30-kb genome in just 10 minutes. To generate large double-stranded DNA amplicons, we exploited repliQa HiFi ToughMix for rapid amplification of longer targets (>10 kb) using an extension time of 10 kb/s. Three overlapping primer sets captured the entire 30-kb genome of OC43. Resulting dsDNA (~10 kb) can be used for long range sequencing with the Oxford Nanopore or PacBio platform. Alternatively, longer amplicons can be used in conventional NGS workflows after DNA fragmentation. Finally, we expanded on this work using SARS CoV-2 RNA (2019-nCoV/USA-WA1/2020) to demonstrate the ability to generate long amplicons available for long-read sequencing.

Overall, longer and faster cDNA synthesis followed by rapid, long-range, high-fidelity PCR enables an expeditious and reliable method to monitor virus evolution, mutation surveillance and pathogen detection.

Hi-Plex RNA spatial phenotyping of human tissues at single-cell resolution

Technology Development

POSTER 549

Julia Kennedy-Darling[1], **Lydia Hernandez**[1], **Will Agnew-Svoboda**[1], **Ming Yu**[1], **Suyog Pathare**[1], **Arturo Gamboa-Aldeco**[1], **Garry P. Nolan**[2]

1 Akoya Biosciences, Menlo Park, CA
2 Stanford University School of Medicine, Stanford, CA

Spatial Biology has the potential to revolutionize our understanding of disease mechanisms and how to develop new therapies. To date, there has been a trade-off between plex-level, resolution, and sample integrity when mapping the contents of a specimen while retaining tissue context. With next-generation sequencing tools, tissue samples are typically dissociated to enable either bulk or single-cell assessment of gene signatures. This information, however, is void of spatial context and is dependent on equivalent dissociation propensities of different cell types for an unbiased view. While there are technologies that enable spatial profiling, but they typically lack single-cell resolution or support low-plex analysis. Further, different markers benefit from measuring the protein product or the RNA product depending on their expression level, localization, and whether quality antibody products exist. Here, we present the new Hi-Plex RNA methodology for analysis of dozens of RNA targets on human tissue samples with single-cell resolution using the CODEX. This technology uses a targeted, amplification-based chemistry to detect markers through cycles of hybridization, imaging, and dehybridization. In this manner, spatial hi-plex RNA data is generated while retaining tissue integrity. Cell types are identified through the specific panels and associated cellular neighborhoods can be detected based on these annotations. This capability complements the existing single-cell, whole-slide high-plex spatial protein methodology that is currently enabled by the CODEX instrument. The combination of this new Hi-plex RNA spatial capability with the existing high-plex protein spatial capability makes the CODEX instrument the first-of-its-kind instrument to enable multiomic analysis of samples at a single-cell resolution platform.

High-throughput single-cell RNA-seq applied to microorganisms

Technology Development

POSTER 550

Julie Laliberte

Singleron Biotechnologies GmbH, Cologne, Germany

Single-cell RNA technologies have rarely been adapted to microorganisms because of challenges related to their size, lower RNA content and cell wall. In 2019, Nadal-Ribelles et al. published the first established single-cell RNA-Seq protocol compatible with the yeast model. Since then, multiple groups showed interest in adapting high throughput single cell technologies to microbial organisms. Here we showed that a single-cell transcriptomic microwell based technology was more flexible and easily adapted than a droplet-based approach. We developed the first commercially available product to enable single-cell RNA-Seq compatibility with yeast models like S. cerevisiae, S. pombe and C. albicans. The chemistry was optimized to capture up to 30,000 single cells per run on a fully automated cell parsing instrument.

Single cell RNA-Seq of isogenic cell populations is currently used to shed light on the transcriptional heterogeneity of microbial processes including sporulation, cell communication, carbon utilization, and competence. Here we used carbon source change to demonstrate the capability of our assay to identify heterogeneity in the repression of GAL1 promoter following glucose exposure. The performance of the assay was demonstrated using S. cerevisiae. The cells were grown in media containing galactose, then transferred to glucose as main source of carbon for a defined number of hours. The cells were immediately fixed with methanol. For single cell RNA-Seq, the cells were rehydrated and counted before being loaded on a chip containing 250,000 microwells. Once on the chip, the yeast's cell wall was digested, and the cells were lysed. The released mRNA molecules were than captured using barcoded beads carrying unique cell barcodes and unique molecular identifiers (UMI). The barcoded beads were then retrieved from the chip and the mRNA molecules were reverse transcribed and amplified. Illumina compatible libraries were obtained by enriching the 3' end of the initial mRNA. The libraries were sequenced on an Illlumina NovaSeq. The data was analyzed using custom tools for cell barcode retrieval, UMI counting, read mapping and annotation.

Between 15,000 to 30,000 cells were captured per sample. Less than 3% of the wells showed multiplets. In average, a total of 6,600 genes were identified per sample. Over 65% of the bases sequenced belonged to exons and were used for annotation and clustering. As expected, genes under GAL1 promoter regulation were repressed in presence of glucose only in a fraction of the cells.

High-resolution multi-omics spatial profiling of whole transcriptome and epigenome in tissue

Technology Development

POSTER 551

Colin Ng[1], **Molly Wetzel**[1], **Jennifer Garbarino**[1], **James McGann**[1], **Joe Gennaro**[1], **Liya Wang**[1], **Ken Wang**[1]

1 AtlasXomics Inc., New Haven, CT

The advent of spatial transcriptomics has enabled researchers to build upon the success of single-cell assays to enhance the understanding of biological systems and pathogenesis. However, cells richly interact with their environment along multiple networks, not just through the transcriptome. Without in situ co-profiling with these other networks, such as the epigenome and proteome, understanding of biological function and disease in structurally complex tissues such as solid tumors and CNS tissues will remain incomplete.

Here we present AtlasXinsights, our service offering transformative spatial maps generated by our discovery platform, Deterministic Barcoding in Tissue for spatial omics sequencing (DBiT-seq). This service will provide comprehensive, unbiased, multi-omics mapping of cell states. Our microfluidics approach features unparalleled compatibility with a variety of tissue types and fixation methods, including FFPE. We demonstrate the multi-omics capacity of AtlasXinsights with results from the first in-tissue, cellular resolution co-profiling of the whole transcriptome (>23,000 genes) and genome-wide chromatin accessibility to identify unique cell-cell interactions underlying biological function in tissue.

This service is the first commercial translation of the DBiT-seq platform, which was licensed exclusively from Yale University. Invented by the research group of Dr. Rong Fan, Professor of Biomedical Engineering at Yale, DBiT-seq's unique microfluidics and Next-Generation Sequencing-based approach enables high-fidelity spatial barcoding of nearly any analyte on a wide variety of tissues. AtlasXinsights will be available to provide the first high-resolution, unbiased, multi-omics mapping of tissue, empowering researchers to decipher the complex networks underlying biological function.

Ultrafast and highly sensitive full-length single-cell RNA-sequencing using FLASH-seq

Technology Development

POSTER 552

Simone Picelli[1,2], **Vincent Hahaut**[1,2], **Dinko Pavlinic**[1,2], **Walter Carbone**[3], **Sven Schuierer**[3], **Guglielmo Roma**[3], **Pierre Balmer**[1,2], **Magdalena Renner**[1,2], **David Rawling**[4], **Joseph Mellor**[4], **Cameron Cowan**[1,2]

1 Institute of Molecular and Clinical Ophthalmology Basel, Basel, Switzerland
2 Department of Ophthalmology, University of Basel, Basel, Switzerland
3 Chemical Biology and Therapeutics, Novartis Institutes for Biomedical Research, Basel, Switzerland
4 seqWell, Beverly, MA

Single-cell RNA-sequencing (scRNA-seq) has undergone exponential growth in the past 10 years. Emulsion droplets methods, such as those commercialized by 10x Genomics, have enabled robust, reproducible analysis of thousands of cells in parallel. However, in contrast to full-length sequencing protocols like Smart-seq 2/3, 10x Genomics technology captures only the 3´-end or the 5´-end of each transcript, thus lacking the coverage and the sensitivity for a comprehensive analysis of the entire transcriptome. Moreover, 3´/5´-sequencing can´t discriminate between splicing isoforms, identify internal SNPs or establish allele usage in different cells.

Building upon the existing SMART-seq 2/3 protocols we developed FLASH-seq (FS), a new full-length scRNA-seq method with increased sensitivity (higher gene detection), limited hands-on time and the potential for customization to meet demands of specific biological questions. Every step in FLASH-seq has been carefully optimized by testing hundreds of conditions to find the best-performing combination of buffers, additives and enzymes. FLASH-seq combines Reverse Transcription (RT) and pre-amplification in a single reaction, decreases RT time by using a more processive enzyme and boosts template switching reaction efficiency by adding an unbalanced mixture of deoxyribonucleotides.

FLASH-seq is an automation-friendly, modular protocol that can be efficiently miniaturized to reduce library preparation costs. By reducing the number of PCR cycles required for cDNA synthesis and eliminating the need for sample cleanup, QC and normalization, we have developed an iteration of the protocol, FLASH-seq Low-Amplification (FS-LA), which enables generation of sequencing-ready libraries in just 4.5 hours, with less than 30 minutes of hands-on time. We have also demonstrated compatibility with the commercially available plexWell library preparation kit (seqWell), which further simplifies the workflow and improves normalization across samples with variable cDNA yields.

(continue to page 194)

Technology Development

POSTER 552

(continued from page 193)

To address the need for molecule counting and isoform reconstruction, we have introduced a Template Switching Oligonucleotide containing Unique Molecular Identifiers (TSO-UMI) into the FLASH-seq system. Our novel TSO-UMI was designed to minimize the largely neglected phenomenon of strand-invasion during RT along with resulting sequencing artefacts. We introduced a 5-bp spacer between the 3 terminal ribonucleotides and the UMI sequence of the TSO, obtaining a significant reduction in the number of strand invasion artefacts compared to Smart-seq3. Such a modification also dramatically increased cDNA yield and, more importantly, resulted in 15% more detected genes.
We are confident that FLASH-seq has the potential to become the tool of choice when looking for an efficient, robust, modular, affordable and automation-friendly full-length scRNA-seq protocol.

KAPA HyperPETE with pre-capture multiplexing in somatic applications for high-performance NGS results and increased sample handling efficiency

Technology Development

POSTER 553

Nitya Furtado[1*], Florence Crary-Dooley[1*], Mingyang Cai[1*], Brian Godwin[1], Liu Xi[1], Jimmy Tu[1], Liesl Joubert[2], Marius Snyman[2], Mariana Fitarelli-Kiehl[3], Spencer Debenport[3], Jingchuan Li[1]

1 Roche Sequencing Solutions, Pleasanton, CA
2 Roche Sequencing Solutions, Cape Town, South Africa
3 Roche Sequencing & Life Science, Wilmington, MA
*co-first authors

KAPA HyperPETE (Primer Extension Target Enrichment) is a novel target enrichment solution newly launched by Roche, which employs primer extension reactions to specifically capture and release target library molecules for sequencing. It offers the fast turnaround time of amplicon-based workflows and the performance of hybridization-based capture. KAPA HyperPETE is a single day workflow that demonstrates high coverage uniformity and detects all major somatic variant types (SNV, InDel, CNV, MSI status, and novel fusion transcript partners).

Pre-capture multiplex pooling is an established approach in capture-based target enrichment methods to decrease the number of sample reaction tubes handled during enrichment steps. While pre-capture multiplex pooling reduces the reagent cost and hands-on time when processing many samples, it is understood that enrichment performance is negatively impacted. These performance losses are usually minimal when detecting heterozygous and homozygous SNPs or InDels while using high quality genomic DNA, but are more problematic when detecting somatic variants of less than 5% allele frequency. This can be compounded by the use of low quality and/or low nucleic acid input amounts commonly derived from oncology samples.

In this study, we optimized pre-capture multiplexing conditions for oncology samples in our enrichment workflow. The results demonstrate sequencing performance metrics and variant calling results comparable to single-plex capture. KAPA HyperPETE with pre-capture multiplexing improves turnaround time, reduces cost, and increases the number of samples that can be processed simultaneously while retaining high performance.

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

Zero-PCR, UMI-independent targeted sequencing for accurate genomic profiling

Technology Development

POSTER 554

Yun Bao

Avida Biomed Inc., Fremont, CA

PCR bias and errors, laborious workflow, limited multiplexing and high reagent cost are the main hurdles of the current targeted sequencing technologies—namely, hybridization- and amplicon-based target enrichment and sequencing.

Here, we report a new targeted sequencing strategy based on the highly efficient chemistry of the Point-n-Seq™ platform called "Zero-PCR". With this method, DNA target molecules are ligated to full-length sequencing adapters and captured without any amplification—before, or after enrichment. A sequencing-ready captured library is prepared with high uniformity for any intended targets in less than 4 hours. By removing all PCR steps from the workflow, the Zero-PCR protocol eliminates the common problems associated with DNA amplification-base pair errors, amplicon bias, GC-bias, and the need for purification and quantification.

Specifically, with Zero-PCR, these problems are solved at once: 1) there are no duplicate reads—every single sequence read is from the original molecule, removing the need for deduplication and significantly reducing sequencing cost; 2) Copy Number Variation (CNV) detection inherently becomes much more accurate due to the elimination of amplification bias; 3) without PCR amplification, the sample capture workflow streamlines the process and enhances the target coverage uniformity, further reducing sequencing cost (demonstrated with a 120kb panel where more than 99% of targets had coverage above the 0.5X mean with a 88% on-target rate); and, 4) without errors introduced by PCR, the need for some complex error correction techniques used in variant detection, such as unique molecular identifiers (UMI), is also greatly reduced.

In addition, we utilized a library construction method to combine DNA fragmentation with end-repair in a single step, enabling all steps to be performed in a single tube, with no cleanup or sample loss. We have demonstrated the feasibility using starting material as 100 ng of intact genomic DNA with panels of 100 kb, or greater. The entire Zero-PCR process with its single-plex procedure (no pooling for capture) creates a seamless workflow with minimized risk of sample mislabeling and contamination. This protocol is currently suitable for genotyping and genomic alteration detection in samples with adequate DNA input (>100 ng) and moderate Limit Of Detection (LOD) requirements (1%).

In brief, the Point-n-Seq technology enabled Zero-PCR targeted sequencing method enhances accuracy and uniformity, simplifies execution, reduces the need for UMI-based error correction, all while lowering overall sequencing cost. This method accelerates the development of fast, simple and accurate genomic profiling assays for current and future clinical applications.

Automated high-yield sequencing library prep for Oxford Nanopore Technologies

Technology Development

POSTER 555

Abdul Mohammed[1], Simran Kaushal[1], Febriana Pangestu[1], Devin Junkins[1], Will Langford[1], Udayan Umapathi[1]

1 Volta Labs, Inc., Cambridge, MA

Here, we present an approach for high-yield sequencing library preparation for Oxford Nanopore Technologies (ONT) using the Volta Labs pipetting-free Digital Fluidics system.

We developed a library prep workflow on the Digital Fluidics system based on ONT's ligation-based kit SQK-LSK-110. The ONT libraries produced by the Digital Fluidics system gave equivalent pore occupancy, read length N50, and mean read lengths when run on a MinION flow cell as manual preps using the same kit. For the same input material, libraries prepared using the Digital Fluidics system resulted in up to 40% higher library yields than manual library preps - the outcome is more sequenceable libraries and an increased number of sequencing reads, resulting in better genomic coverage.

The Digital Fluidics system was also able to process kilobase length to megabase lengths enzymatically (for repair and adapter ligation) and perform bead-based manipulation. Sequencing the molecules processed using this approach gave 2X higher read length N50s for megabase length fragments compared to hundred kilobase DNA. Not only does the Digital Fluidics system allow for efficient conversion during library prep for long-read sequencing, but this demonstration shows the suitability of this approach for handling a high dynamic range of input DNA fragment lengths. The implication of our preliminary data is that users can handle a variety of sample inputs, will be able to get more data per input (or reduced cost per base), and will have the flexibility to start with less input material to achieve acceptable data quality from libraries prepared by our system.

Targeted sequencing of a 1 Mb expanded-carrier screening panel using molecular inversion probes

Technology Development

POSTER 556

Patrick C. Saunders[1], **Jack M. Amaral**[1], **Gregory J. Porreca**[1], **Eric D. Boyden**[1]

1 Molecular Loop Biosciences, Inc., Alltson, MA

Target capture for next generation sequencing (NGS) is typically performed using either of two methods: hybridization capture, which produces high quality data but suffers from a complicated and labor-intensive workflow, and PCR capture, which has a simpler workflow but may be prone to poor uniformity and allele dropout. We have developed and optimized an alternative method based on molecular inversion probes (MIPs) that combines the robust data quality of hybridization capture with a workflow that is even simpler than that of PCR capture. The superior performance of our MIP capture protocol is enabled by dense bi-stranded probe tiling that ensures high coverage uniformity and resistance to allele dropout, while benefitting from a 4-step addition-only workflow that eliminates the creation of a shotgun library and all sample-level purifications.

Although MIPs are often used to capture relatively small target regions (e.g. <100 kb or up to dozens of genes) with modest numbers of probes, we hypothesized that there is no theoretical limit on the overall target size or number of probes. Accordingly we developed a panel of MIPs designed to capture over 1 Mb of genomic targets, composed of coding sequences, splice sites, and 5' UTRs for 339 genes that are commonly targeted in expanded carrier sequencing (ECS) panels. The panel consists of >40,000 MIPs tiled against both DNA strands to achieve an average of 5 probes per target base. Each MIP captures an element that is nominally 225 bp, which is ideal for Illumina PE150 sequencing and optionally enables stitching of overlapping paired reads into longer single reads for improved sensitivity to indels. Sequencing adapters and unique dual indices (UDIs) are incorporated into the PCR primers, which allows multiplexing of hundreds to thousands of samples without risk of index cross-talk. Our protocol is highly scalable, easily automatable, requires less than 2 hours of hands-on labor, and may be performed using either a single-day or overnight sample-to-sequencer workflow.

This work demonstrates that the target size for MIP capture is not technologically limited. Even a full exome panel may be possible that would substantially improve on the simplicity and scalability of existing hybridization-based exome capture protocols without compromising clinical-grade data quality.

High-resolution multiomic spatial phenotyping: a union of CODEX and DBiT-seq

Technology Development

POSTER 557

Oliver Braubach[1], **Hailing Zong**[1], **Matt Eastman**[2], **Bassem Ben Cheikh**[1], **Colin Ng**[2]

1 Akoya Biosciences, Menlo Park, CA
2 AtlasXomics, New Haven, CT

Spatial biology has revolutionized the life sciences by providing ultra-high parameter imaging of protein and RNA in intact tissues. Progress has been swift on either side, and it is now possible to detect dozens of proteins or RNA transcripts within single tissues at single cell resolution. The resulting datasets have started to dominate discovery to translational research agendas and are expected to influence clinical research. Yet, progress towards creating multiomic spatial biology solutions has been slow, and there is no single technique that combines ultra-high parameter imaging of both protein and RNA in situ and at good spatial resolution.

Akoya Biosciences' CODEX platform enables the measurement of dozens of proteins with spatial context, across whole slides, and at single cell resolution. AtlasXomics' DBiT-seq is an in situ spatial transcriptome mapping technique that allows unbiased whole transcriptome profiling (>24,000 genes) at high resolution. To enable spatial multiomics at high resolution from the same tissue, we developed a combinatorial workflow. Tissues are serially processed using DBIT-seq and CODEX and imaged using Akoya's CODEX platform; the fastest, single-cell, whole-slide spatial phenotyping solution. DBiT-seq and CODEX employ mechanistically similar microfluidic strategies to extract and label analytes, respectively, and neither destroys the tissue. Both analyses are done on the same tissue, and it is thus possible to produce assertive multiomic results in FFPE and fresh frozen samples. We present data that show high concordance of mRNA and protein expression with hitherto unmatched sensitivity and resolution in human FFPE and mouse fresh-frozen brain tissues. The combined workflow is immediately deployable and collectively offers the highest analyte and spatial resolution among competing platforms.

A high-sensitivity 'liquid biopsy' assay using Linked Target Capture (LTC) technology

Technology Development

POSTER 558

W. Lloyd Ung[1], **Kathleen Kolehmainen**[1], **David Broemeling**[1], **Andre Marziali**[1,2], **Joel Pel**[1]

1 Boreal Genomics Inc., Vancouver, British Columbia, Canada
2 Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia, Canada

The non-invasive analysis of circulating tumour DNA (ctDNA) from blood via a 'liquid biopsy' is rapidly emerging as a powerful tool in the detection, monitoring and treatment of cancer. In particular, targeted Next Generation Sequencing (NGS) has enabled the detection of many rare ctDNA cancer biomarkers in a single assay. However, existing target enrichment technologies present limitations. Multiplexed PCR methods are rapid but struggle to produce high uniformity over large panels (limiting detection of variants such as CNVs), cannot easily deliver duplex UMI information (or template start/stop positions), and often require sample splitting across multiple primer pools, reducing sensitivity. Hybrid capture methods, on the other hand, deliver low on-target rates for small panels, generally require higher input mass, have long and complex workflows, and lack the specificity of amplicon methods.

Linked Target Capture (LTC) technology, which uses physically linked capture probes and PCR primers, has been optimized to enable the best features of both multiplex PCR and hybrid capture in a single liquid biopsy assay. The LTC workflow can be completed in a single work day, requires minimal hands on time, and because it uses a linked probe-pair capture approach, has high sensitivity and specificity for desired sequences. LTC also enables the detection of gene fusions with unknown partners, since the linked probe-primer pair can capture across an untargeted sequence. Additionally, LTC is compatible with single or duplex UMIs, preserves template start/stop locations, produces high uniformity, and can deliver high on-target regardless of panel size. A machine learning design pipeline has been shown to generate panels that meet stringent performance requirements without re-design, making custom panel applications possible.

We have developed a 52-gene 'pan-cancer' liquid biopsy panel using LTC. In addition to including the most common SNV and indel biomarkers, the panel enables full coverage of TP53 without sample splitting, the detection of gene fusions with known and unknown partners, and the detection common CNV genes. We demonstrate panel performance, including high uniformity and on-target, low error rate, single-molecule sensitivity, and detection of rare variants, including gene fusions, below 0.1%.

FocuSCOPE: a single-cell, multi-omics solution to simultaneously analyze tumor variants and microenvironment

Technology Development

POSTER 559

Wenqi Zhu, **Yiqi Zhou**, **Nan Fang**

Singleron Biotechnologies, Köln, Germany

Recent advances of high-throughput single cell sequencing technologies have greatly improved our understanding of the complex biological systems. Heterogeneous samples such as tumor tissues commonly harbor cancer cell-specific genetic variants and gene expression profiles, both of which have been shown to be related to the mechanisms of disease development, progression, and responses to treatment. Furthermore, stromal and immune cells within tumor microenvironment interact with cancer cells to play important roles in tumor responses to systematic therapy such as immunotherapy or cell therapy. However, most current high-throughput single cell sequencing methods detect only gene expression levels or epigenetics events such as chromatin conformation. The information on important genetic variants including mutation or fusion is not captured. To better understand the mechanisms of tumor responses to systematic therapy, it is essential to decipher the connection between genotype and gene expression patterns of both tumor cells and cells in the tumor microenvironment. We developed FocuSCOPE, a high-throughput multi-omics sequencing solution that can detect both genetic variants and transcriptome from same single cells. FocuSCOPE has been used to successfully perform single cell analysis of both gene expression profiles and point mutations, fusion genes, or intracellular viral sequences from thousands of cells simultaneously, delivering comprehensive insights of tumor and immune cells in tumor microenvironment at single cell resolution.

A multimodal, label-free tool for gentle and efficient cell separation

Technology Development

POSTER 560

Kevin Travers[1], **Seth Brodie**[1], **Susana Jett**[1], **Chris Stevens**[1], **Suzanne Weaver**[1], **Teddy de Groot**[1], **Stefan Ha**[1], **Geoff Facer**[1]

1 LevitasBio, Inc., Menlo Park, CA

LevitasBio has developed the LeviCellTM platform, a powerful new solution for cell processing and characterization that utilizes magnetic fields to levitate cells. Unlike other single cell separation and cellular analysis methods, levitation does not require dyes, antibodies or specific markers, and the cells are not modified or perturbed in any fashion. In this novel cell separation platform, an inert, paramagnetic compound is added to the media in which the cells are suspended. The cells are added to a single-use cartridge, which is then inserted into the LeviCell where the cells equilibrate in the presence of an externally applied magnetic field. The cells levitate in solution to specific heights determined by the cells' intrinsic properties, including density and magnetic susceptibility. The cell suspensions are imaged in real-time, enabling complete analysis and characterization of the sample with real-time control over the cell separation. The simplicity of the system allows cells to be treated gently without use of high pressure or other perturbations that commonly lead to increased cellular stress responses, cell activation, or cell death. The simple, direct flow path leads to high yields of cells even with very low input cell numbers.

This novel new technology has been applied in a variety of applications, including the enrichment of viable cells from dead or dying cells and debris, separation of different cell types possessing distinct densities, and the separation of differentiated from undifferentiated cells. Here we demonstrate the performance of the LeviCell platform in the enrichment of viable cells from a variety of samples, including cell lines and primary samples such as PBMCs and dissociated tumor samples. Dissociated tumor samples with a starting viability of 50-60% are readily enriched to viabilities of greater than 85%. Gene expression studies comparing cells pre- and post-separation highlight the gentleness of the enrichment approach. The ability to enrich live cells with high recovery is particularly significant for analytical applications such as single cell sequencing, as well as cellular expansion workflows such as cell line development, iPSC workflows, and CRISPR recovery. Here, we demonstrate significant improvements in the number of cells and transcripts identified in single cell sequencing following enrichment of viable cells on the LeviCell.

High-throughput library preparation for HiFi whole genome sequencing

Technology Development

POSTER 561

Aaron Wenger, **Heather Ferrao, Christine Lambert, Vicky Liu, Lin Wang, Joan Wilson, Shreya Chakraborty, Davy Lee, Aurelie Souppe, William Rowell, Janet Ziegle, David Miller, Primo Baybayan, Suzanne Dee, Aurelie Souppe**

Pacific Biosciences (PacBio), Menlo Park, CA

HiFi reads combine long read lengths (>10 kb) and high accuracy (99.9%), which has enabled the complete and accurate characterization of whole genomes, including the first telomere-to-telomere assembly of a human genome. HiFi reads have simplified many bioinformatic analyses – efficient algorithms have been developed for read mapping, variant calling, and de novo genome assembly with HiFi reads. But obstacles have remained in preparing libraries, namely high DNA input requirements (5 µg per human genome) and a multi-step workflow requiring about 12 hours.

We have developed a HiFi library preparation workflow that reduces DNA input requirements, workflow steps, and end-to-end time. Steps have been combined, reagents provided as master mixes, and reaction times optimized. Size selection is performed with paramagnetic beads, which reduces workflow time and improves DNA recovery compared to size selection with gel cassettes. In total, the new workflow – based on PacBio SMRTbell Prep Kit 3.0 – reduces 10 steps to 7 and decreases workflow time from 12 hours to 6 hours.

We applied the workflow to generate libraries for a variety of samples including individual humans and multiplexed microbe strains. All samples produced high-quality libraries that resulted in high HiFi sequencing yields and excellent analysis results. The human sample HG002 achieved accurate variant calling (>99.9% SNV F1, >99.5% indel F1) and contiguous and accurate genome assembly (contig N50 >90 Mb, quality >Q50) at 32× coverage. All microbes assembled into complete genomes, with more than 375 Mb of total genome assembled from a single SMRT Cell.

This improved workflow supports high-throughput generation of libraries for PacBio HiFi sequencing including using liquid handling robots, which is required for rare disease and population genetics projects that are applying HiFi sequencing at new scales.

Development of a novel TIMS-QTOF mass spectrometer for single-cell proteomics

Technology Development

POSTER 562

Gary Kruppa[1], **Kristina Marx**[1], **Renata Blatnik**[1], **Verena T**[1], **Markus L**[1], **Oliver R**[1], **Nagarjuna N**[1]

1 Bruker Daltonik GmbH, Bremen, Germany

Introduction

Single cell proteomics is a young niche of proteomics compared to single cell genomics. In recent years, significant progress has been made toward single cell proteomics through improvements in sample handling, and boosting the sample signal by multiplexing, but coupled to existing mass spectrometry technology. The timsTOF SCP is the first commercially introduced mass spectrometer dedicated for single cell proteomics driven by a substantial boost in sensitivity. The modified front-end of the instrument increases the sensitivity by 4-5 fold ion transfer into the mass spectrometer. Here we demonstrate the performance of the instrument for low sample loads, equivalent to a single cell in combination with robust low flow rate nanoflow LC sample introduction.

Methods

Commercially available K562 tryptic peptide digests (Promega) were onto the LC sample introduction system. Peptide amounts of 250 pg, 500 pg, 1 ng and 10 ng were used to evaluate the performance of the instrument. A low flow method (100 nl/min) and a gradient time of about 28 minutes was used for separating the eluting peptides on a 15 cm 75 micron ID column coupled to captive spray ionization source using a 10-micron ID zero-dead volume emitter. Samples were analyzed using a data independent analysis method (dia-PASEF) tuned for high sensitivity measurements. Raw data were processed in Spectronaut 15 (Biognosys) or DiaNN.

Preliminary data

Using the combination of low flow nanoLC and the high sensitivity data acquisition methods on the timsTOF SCP, we could quantify above 1000 protein groups from 250 pg sample loads and close to 2000 protein groups from 500 pg loads. Performing a dilution series experiments, more than 4000 protein groups could be quantified in 6.4 ng peptide loads. We plan to further increase the depth of proteome quantification by increasing the depth of the library used in the data processing step. Further we plan to investigate the stability and the effect of sample storage of such low sample amounts on the depth of proteome quantification.

Optimization of DNA library preparation workflows for different sequencing platforms

Technology Development

POSTER 563

Pingfang Liu, **Chen Song, Matthew Campbell, Keerthana Krishnan, Bradley W. Langhorst, Eileen T. Dimalanta, Theodore B. Davis**

New England Biolabs, Ipswich, MA

Advances in sequencing instrumentation and associated chemistries continue to evolve, and new sequencers are being released at an unprecedented pace by both the current market leader (Illumina) and market challengers, including MGI. Lower run costs, faster run times and ease-of-use offered by new sequencing companies often makes switching platforms an appealing and reasonable choice. However, data quality, error rates, and error types can vary among sequencers, even between instruments from the same manufacturer. Therefore, it is important to understand the performance of different sequencing platforms to ensure continuity of long-term projects and if necessary, to develop mitigation strategies for error correction.

Here, we present two sample preparation workflows that have been optimized for MGI sequencing platforms. One method utilizes sheared DNA as the starting material, whereas the other method utilizes a robust, flexible, enzymatic fragmentation method; both workflows are compatible with MGI's standard MPS and CoolMPS sequencing chemistries. Using these workflows, we generated high quality data with superior sequencing metrics, such as increased mapping rate, lower chimeras, lower base bias, and increased genome coverage compared to other available workflows. Additionally, our method can be used to generate libraries for multiple sequencing platforms, allowing for direct comparisons of different sequencing technologies.

Rapid whole genome sequencing (rWGS) on a novel digital microfluidic system: "on demand" personal automation for diagnosis of critically ill infants

Technology Development

POSTER 565

Kate Cunningham[1], **Yan Ding**[2], **Julia Yoo**[1], **Sergey Batalov**[2], **Alma Johnson**[2], **Eduardo Cervantes**[1], **Adam Barner**[1], **Eugenia Carvajal**[1], **Cheng-Chang Lee**[1], **Severine Margeridon**[1], **Mais Jebrail**[1], **Wendy Benson**[2], **Foteini Christodoulou**[1], **Stephen Kingsmore**[2]

1 Miroculus, San Francisco, CA
2 Rady Children's Institute for Genomic Medicine, Rady Children's Hospital, San Diego, CA

Genetic disorders and congenital anomalies are a leading cause of morbidity and mortality in infants, affecting ~2% of live births in the United States. There is strong evidence that rapid, clinical Whole Genome Sequencing (rWGS) at presentation of critically ill infants enables precision medicine interventions for many of these genetic diseases that improve outcomes and reduce cost of care. There is now a need to scale rWGS for national adoption. However, manual rWGS library preparation is laborious and lengthy. Critically ill infants cannot afford to wait for sample batching on traditional high throughput automation solutions, and could benefit from a system that consistently generates high quality DNA libraries for rWGS with minimal manual effort. Here, we report a novel solution: the digital microfluidic (DMF)-based system, Miro Canvas. Miro Canvas utilizes electromechanical forces to move, merge, mix, and dispense microliter-volume fluids across a surface of patterned electrodes in an automated fashion. Inter-sample contamination is prevented by constructing libraries within single-use, electronics-free cartridges.

We automated the Illumina DNA PCR-free Tagmentation Prep Kit for rWGS library preparation on the Miro Canvas and compared the library quality with manual and high-throughput liquid handler-based methods. The input was 500ng purified genomic DNA samples from reference materials (Coriell Repository NA12878) and patient samples and their family members (trios) with research consent at the Rady Children's Institute for Genomic Medicine. Library quantity was assessed by qPCR or ssDNA Qubit and WGS was performed using NovaSeq6000 systems. Library yield, insert size, genome coverage, OMIM gene coverage, and variant calling sensitivity/specificity was comparable to that of standard methods. Our preliminary results showed that the Miro Canvas can generate comparable libraries for downstream rWGS that provide high confidence variant calling, including structural variants. Analytic sensitivity, specificity and positive predictive value of SNP (single nucleotide polymorphisms) and indel (insertion and deletion) were higher than 99%. The variants of interest identified previously through a clinically validated rWGS workflow were recaptured 100% in the selected samples by rWGS through this library preparation. On average, hands-on time and turnaround time (TAT) of library preparation on the system were approximately 40 minutes and 110 minutes, respectively.

In conclusion, a DMF-based system generated high-quality rWGS libraries with minimal manual effort. For singleton, duo, and trio samples from critically ill children, this system provides a novel solution for automated library preparation when waiting for centralized or batched sample processing is not a viable option.

Highly specific single-cell spatial transcriptomics in FFPE enabled via in situ probe ligation-based assays

Technology Development

POSTER 566

H. Benjamin Larman[1,2], **Joel J. Credle**[1,2], **Boyoung Cha**[2], **Sun Y. Lee**[2], **Chloe Kim**[3], **Dan Xie**[3], **Marc Glazer**[3], **James Dietz**[3]

1 Department of Pathology, Institute for Cell Engineering, Johns Hopkins University, Baltimore, MD
2 Portal Bioscience, Baltimore, MD
3 Veranome Biosystems, Mountain View, CA

Archival tissues from surgical resections and biopsies constitute a largely untapped resource for biomedical research. However, the typical tissue preservation procedure, formalin fixation and paraffin embedding (FFPE), causes RNA degradation due to strand fragmentation, crosslinking, and adduct formation. We have previously reported the development of a probe ligation-based method for quantifying RNA in FFPE sections called Ligation In Situ Hybridization sequencing (LISH-seq) which avoids reverse transcription and RNA extraction, steps that limit traditional analyses of FFPE RNA. LISH-seq is an extremely efficient method for measuring the abundances of hundreds to thousands of mRNA targets in even very small FFPE regions of interest collected from single tissue sections. Here, we show how amplified products formed in situ using probe ligation followed by rolling circle amplification can be detected via hybridization with fluorescently labeled oligonucleotides. We refer to this new system of spatially resolved RNA analysis as LISH-Lock'n'Roll. The technology can be used to measure fixed mRNA expression levels and splicing, as well as fusion products and expressed mutations or single nucleotide polymorphisms, with single cell spatial resolution. Using Veranome Biosystems's automated fluidics and imaging instrument, VSA-1, LISH-Lock'n'Roll can be used for single cell analysis of >100 targets per section. As a demonstration of their integrated utility, LISH-seq and LISH-Lock'n'Roll were together used to investigate the connection between inflammatory bowel disease (IBD) and the elevated risk for developing colorectal cancer (CRC). Over 2,000 LISH-seq probes were used to interrogate FFPE surveillance biopsies taken from individuals that did or did not develop CRC, who either did or did not have pre-existing IBD. A set of LISH-seq probes targeting differentially expressed mRNAs were converted into spatial LISH-Lock'n'Roll probes and used for single cell analyses. This study illustrates the utility of LISH-based assays for analysis of gene expression in FFPE, both in regions of interest and at single cell resolution.

Long-read assay development for Illumina sequencers

Technology Development

POSTER 567

Gary Schroth, Feng Chen, Mike Eberle, Aaron Halpern, Ryan Lusk, Rami Mehio, Steve Bruinsma, Joshua Burgess, Aaron Darling, Haiying Grunenwald

Illumina Inc., Foster City, CA and San Diego, CA

Illumina is developing a novel long read assay that includes both assay and software components and is fully compatible with NovaSeq sequencing system. The streamlined and automation enabled workflow requires as little as 50 ng of input genomic DNA. The dedicated DRAGEN pipeline software processes standard 2x150 bp paired-end data to create > 5 kb contigs. Combining the long read assay and Illumina's standard PCR-Free 30X human data, we are able to deliver a higher resolution view of the genome with improved capability to capture complex variants, including previously difficult to sequence regions. Such combination provides the most complete view of the human genome available from Illumina platforms – providing exceptional coverage and mapping of complex structural variants and highly repetitive regions of the genome. We will demonstrate examples of discrete regions throughout the genome where the combination of short and synthetic long reads provides improved mapping and variant calling. In addition, an enrichment workflow leveraging this innovative technology can capture regions where short reads alone have some ambiguity which will further extend its application to support targeted clinical solutions for oncology and genetic diseases.

Efficient whole-genome methylation sequencing using mostly natural SBS reveals evolution of methylation patterns associated with the development of colorectal cancer

Technology Development

POSTER 568

Zohar Shipony[1], Gil Hornung[1], Sarah Pollock[1], Tyson Clark[1], Florian Oberstrass[1], Doron Lipson[1], Hayan Lee[2], Aaron Horning[2], Graham Erwin[2], Michael Snyder[2]

1 Ultima Genomics, Newark, CA
2 Department of Genetics, Stanford University School of Medicine, Stanford, CA

Familial adenomatous polyposis (FAP) is an inherited disorder characterized by predisposition to develop colorectal cancer (CRC) and numerous colorectal adenomatous polyps. Many studies looked at the numerous changes that occur in DNA methylation during CRC progression, but because of the high cost of sequencing they were limited to few samples at full genomic coverage or many samples looking at only an enriched fraction of genomic CpGs.

Mostly-natural SBS (mnSBS) is a new sequencing technology that enables low-cost, high-throughput sequencing via a combination of non-terminating chemistry and large-scale optical scanning. It is specifically advantageous for whole genome methylation sequencing (WGMS) since the reduced complexity of a converted genome allows generation of single-end reads of length >250bp, which improves their mappability in a whole-genome context. We first tested the performance of WGMS using mnSBS by sequencing EM-seq converted libraries generated from the reference cell lines GM12878 and K562, demonstrating >0.94 concordance vs existing ENCODE data for % methylation over 23M common CpGs.

Using mnSBS we were able to produce high-coverage and high-quality WGMS data for 18 EM-seq samples from 4 different individuals on normal mucosa, dysplasia polyps, and adenocarcinoma from early CRC patients, with >55X mean coverage and 240bp mean read length covering 10 different polyps, 5 normal/blood/germline and 3 CRC samples. We observe thousands of DNA methylation changes between normal, polyps, and cancerous samples with a gradual loss or gain of DNA methylation at different genomic contexts. Many of the gradual methylation changes occur at CpG islands of colon development-related genes, including many genes that are known to change in expression during CRC progression. Given the high CpG coverage (over 25M CpGs covered at >10X) we observe substantial change between normal and cancerous cells or adenocarcinoma at hundreds of different enhancers with strong enrichment for Krüppel-Like Factors (KLF) family of transcription factors. The availability of long reads covering multiple CpGs with high coverage allows us to study the sub-population structure and its variation during tumor progression. Combining high-quality whole genome sequencing (WGS) with WGMS allows for detection of genomic regions where SNVs are correlated with methylation changes at different epigenetic states.

Using mnSBS technology for WGMS can help move DNA methylation from the era of arrays to whole genome measurements that provide access to all ~30M CpG sites in the human genome, enabling new discovery and better understanding of DNA methylation in both normal and disease tissues.

Profiling DNA methylation at single-nucleotide resolution in plasma cell-free DNA using whole-genome bisulfite sequencing (WGBS)

Technology Development

POSTER 570

Yi Xu[1], **Caila Ruiz**[1], **Hanjun Kim**[1], **Zhenguo Zhang**[1], **Jeffrey Bhasin**[1], **Mingda Jin**[1], **Zhenfeng Liu**[1], **Angela Garibaldi**[1], **Larry Jia**[1], **Xiaojing Yang**[1]

1 Zymo Research Corporation, Irvine, CA
Corresponding author's email address: yxu@zymoresearch.com

Circulating cell-free DNA (cfDNA) has been reported to contain valuable genetic and epigenetic information for the diagnosis and prognosis of cancer. Studies have shown that the blood from cancer patients contained more tumor-derived cfDNA compared to the healthy controls. DNA methylation, the most well-studied epigenetic marker, has been validated as one of the key drivers in the development of many diseases including cancer. cfDNA methylation thus holds great potential to become a biomarker that will enable early detection of cancer. However, cfDNA is also a challenging input for library preparation as cfDNA is mostly highly fragmented, and the amount of cfDNA from blood plasma is comparatively low. Therefore, multiple library preparation protocols were evaluated for profiling DNA methylation patterns at single nucleotide resolution from ultra-low amount of blood plasma cfDNA. These methods included two bisulfite conversion-based protocols and an enzymatic based protocol. The bisulfite conversion reaction was optimized allowing a milder treatment for less nucleic acid damage. One of the bisulfite conversion based protocols also contained a further optimized adapter ligation step, allowing a simpler procedure. Circulating cfDNA were extracted from blood plasma of a healthy control, a non-small cell lung cancer patient, and an adenocarcinoma lung cancer patient. 5-ng cfDNA was used as input and the libraries were successfully prepared using all three methods. Each library was sequenced to 300 – 600 million read pairs at a read length of 150 bp, enabling a mean coverage of about 10X per detected CpG, and over 90% of the CpG sites in human genome were detected. Unique alignment rate among all libraries were about or above 80%. The comprehensive coverage of the CpG sites across the entire human genome illustrated that the optimized bisulfite conversion reaction was compatible with fragmented cfDNA, proving the satisfactory efficacy in this simple and classic method for DNA methylation profiling of such a challenging sample type. This opens the door to apply the streamlined and simple Whole-Genome Bisulfite Sequencing (WGBS) protocol for cfDNA methylation profiling, which in turn shall facilitate further understanding in cfDNA epigenetic variations and advance the development of a cancer biomarker from the non-invasive, liquid biopsies.

Purifying extrachromosomal DNA from cancer cells for targeted analyses and amplicon reconstruction using CRISPR-CATCH

Technology Development

POSTER 571

King L. Hung[1], **Siavash R. Dehkordi**[2], **Jens Luebeck**[2,3], **Vineet Bafna**[3], **Howard Y. Chang**[1,4]

1 Center for Personal Dynamic Regulomes, Stanford University School of Medicine, Stanford, CA
2 Bioinformatics and Systems Biology Graduate Program, University of California at San Diego, La Jolla, CA
3 Department of Computer Science and Engineering, University of California at San Diego, La Jolla, CA
4 Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA

Oncogene amplification is a key driving mechanism of cancer. Oncogenes are frequently amplified on extrachromosomal DNA (ecDNA), accounting for nearly half of human cancers and 14% of tumor samples. It has been suggested that ecDNA contains genetic alterations including single-nucleotide polymorphisms (SNPs) and structural variants (SVs) that can be distinct from chromosomal DNA. ecDNA also contains more accessible chromatin, implying that it may be epigenetically distinct. A method for purifying ecDNA molecules without amplification would allow us to better identify these features. Using in-vitro CRISPR-Cas9 digestion followed by pulsed field gel electrophoresis, we were able to purify ecDNA from cancer cells. Using a gel extraction protocol similar to that previously described for bacterial DNA from pulsed field electrophoresis (termed CRISPR-CATCH), we confirmed strong enrichment of ecDNA from a panel of cancer cell lines. We identified a number of oncogene SNPs and SVs that are uniquely present on ecDNA in a glioblastoma sample, suggesting that these genetic alterations may have been selected on ecDNA. As DNA amplification is not required, CRISPR-CATCH allowed us to analyze epigenetic features of purified ecDNA molecules such as CpG methylation. Finally, we developed an analytical pipeline for reconstructing complex ecDNA amplicon structures with high confidence, utilizing sequence segment multiplicity and SVs identified from purified megabase-sized ecDNA molecules. The ability to purify and analyze megabase-sized ecDNA molecules in cancer cells will allow us to better identify unique genetic and epigenetic features of ecDNA oncogene amplifications and better understand how they drive cancer.

Linking genome sequence to regulatory function across tissues using an in vivo massively parallel reporter assay and machine learning

Technology Development

POSTER 572

Ashley R. Brown[1,2], **Easwaran Ramamurthy**[1,2], **Irene M. Kaplow**[1,2], **Grant Fox**[1,2], **Morgan E. Wirthlin**[1,2], **Noelle Toong**[1,2], **Ziheng Chen**[1,2], **BaDoi N. Phan**[1,2], **Andreas R. Pfenning**[1,2]

1 Department of Computational Biology, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA
2 Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA

Genetic studies are rapidly identifying an increasing number of human disease-associated loci. Understanding the regulatory mechanisms underlying these loci remains a challenge. It is often unclear which specific candidate genetic variants are causal and which tissues they act in. Massively parallel reporter assays (MPRAs) have the potential to link differences in genome sequence, including disease-associated single nucleotide polymorphisms, to differences in tissue-specific regulatory function. In these experiments, regulatory activity is quantified by having thousands of synthesized regulatory elements control the expression of a corresponding set of unique RNA barcodes, which are measured through RNA-Seq. Although MPRA and similar technologies have been widely adopted in vitro, there have been several barriers to widespread use in behaving animals. In particular, the power of MPRA is limited by the ability to transduce cells, which is more difficult in vivo. To overcome these challenges, we use a new viral strategy and machine learning approach. Experimentally, we deliver the plasmid library using AAV-PHP.eb, which effectively transduces tissues across the mouse, including the liver, kidney, heart, and large portions of the brain. Using this approach, we observe tissue-specific regulatory effects of disrupting transcription factor binding sites and individual nucleotides. To more effectively use MPRA datasets to draw inferences about sequences that are not directly measured, we develop a transfer learning approach. Convolutional neural network regression models linking enhancer sequence to tissue-specific regulation are first trained on open chromatin datasets, which provide more training examples, but are substantially noisier. Then, the models are refined by training on the results of MPRA experiments. Applying transfer learning and including open chromatin leads to a similar performance in predicting regulatory activity from genome sequence when only 25% of the enhancers are used in training (Pearson R=0.39 vs R=0.40). In summary, by improving the transduction efficiency across the mouse with AAV-PHP.eb and by leveraging open chromatin datasets with transfer learning, we can test and interpret the consequence of genetic differences in behaving animals.

A novel enzymatic fragmentation library preparation chemistry that greatly reduces sequencing artifacts and enables the construction of high-quality NGS libraries from an extremely wide range of DNA input quantities

Technology Development

POSTER 573

Martin Ranik[1], **Zane Jaafar**[1], **Josh Haimes**[1], **Thomas D. Harrison**[1], **Lindsay Peterkin**[1], **Kristin Scott**[2], **Daniel Enriquez**[3], **Jonathan Keats**[3], **Brian Kudlow**[1]

1 Watchmaker Genomics, Boulder, CO
2 Colorado State University, Department of Cell & Molecular Biology, Fort Collins, CO
3 Translational Genomics Research Institute (TGen), Phoenix, AZ

Short read NGS library preparation methods which employ enzymatic fragmentation of DNA offer throughput, workflow, and conversion efficiency improvements compared to techniques that utilize DNA fragmented by sonication. However, commercially available enzymatic fragmentation chemistries have been shown to suffer from several undesirable characteristics including decreased coverage uniformity, DNA sequence context bias, sensitivity to DNA input quantity as well as quality, and, most importantly, the tendency to generate artifactual sequences which can lead to false variant calls. We have developed a novel enzymatic fragmentation-based library preparation technology which, together with our high-fidelity library amplification module, effectively overcomes many of the key limitations of similar chemistries. Using whole human genome sequencing as a model, we demonstrate that our methodology offers highly tunable and reproducible library fragment distributions and consistently high conversion efficiencies in a streamlined and automatable configuration. We compared the performance of our kit to leading commercial enzymatic fragmentation library preparation chemistries as well as to techniques that utilize sonicated DNA, and we observed consistently even coverage uniformity across the genome and low sequence-specific bias. Libraries prepared using our chemistry exhibited greatly decreased the incidence of chimeric reads and terminal hairpin artifacts, both of which were observed at rates comparable to those seen in libraries prepared from sonicated DNA and greatly reduced compared to libraries prepared with similar enzymatic fragmentation techniques. The robustness and tunability of the novel chemistry was further interrogated by testing its performance with input amounts below one nanogram. Using a slightly modified workflow, reproducibly fragmented libraries were constructed from as little as one picogram of DNA with minimal residual adapter dimer. Libraries generated from low picogram quantities of input exhibited similar data quality characteristics as libraries constructed from inputs several orders of magnitude higher. The utility of this high dynamic range was demonstrated by assessing the ability to detect known copy number variations (CNVs) in libraries prepared from extremely low DNA amounts without conventional whole genome amplification prior to fragmentation and adapter ligation.

Shedding light on long-range interactions of the human epigenome using ultra-long nanopore sequencing

Technology Development

POSTER 574

Roham Razaghi[1], **Paul W. Hook**[1], **Duncan Kilburn**[2], **Kelvin Liu**[2], **Miten Jain**[3], **Kasper D. Hansen**[1], **Winston Timp**[1]

1 Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD
2 Circulomics Inc., Baltimore, MD
3 University of California, Santa Cruz, CA

Nanopore sequencing has enormous potential in epigenetic applications; unlike traditional sequencing-by-synthesis technologies, it can distinguish covalently modified nucleotides directly through their modulation of the electrolytic current. We can take advantage of the long read lengths (>100kb) generated by nanopore sequencing to precisely call methylation patterns and obtain phased methylation information across the genome. We have used a new high molecular weight DNA extraction technique to generate ultra-long nanopore sequencing data. Using this approach, we can routinely generate over 100Gb sequencing data per PromethION flowcell with an N50 of over 100kb. We generated data on a colorectal cancer cell line (HCT116), non-tumorigenic epithelial cell line (MCF10A), lymphoblastoid cell line (GM12878), and human normal peripheral blood mononuclear cells (PBMC), in order to gain insights into gene regulation measured on single molecules. Gene regulation in mammalian cells is controlled in part by DNA methylation and the complex interplay between methylation at the gene promoter, gene body, distal enhancers, and other regulatory elements. To better understand these long-range epigenetic interactions using ultra-long molecules, we examined phased long-range enhancer-promoter interactions, showing that ultra-long reads can cover up to 215 enhancers/promoters on a single-read level. Through this, we show that we can quantify DNA methylation heterogeneity at interacting regulatory regions. Finally, using cell-type deconvolution approaches, we can use methylation to split the reads in a heterogeneous mixture of cells according to methylation status at known differentiation state biomarkers. Using this we can estimate the proportional cell compositions in synthetic mixtures as well as human blood data.

Single nuclei RNA (snRNA) sequencing: applications in transcriptome and epigenome profiling

Technology Development

POSTER 575

Neha Gujaria-Verma[1], **Gavin Washburn**[1], **Derek Vargas**[1], **Nitin Mandloi**[1], **Kushal Suryamohan**[1], **Chaitanya R. Ekkirala**[1]

1 MedGenome Inc., Foster City, CA

Single-cell RNA sequencing (scRNA-seq) has revolutionized genomics and is now an integral part of therapeutics and diagnostics research. While this technology has enabled the analysis of gene expression at single cell resolution in a variety of research areas including the study of cellular development, the identification of cell types in healthy and diseased states, there are several scenarios where it is difficult to obtain intact, viable cells, a fundamental requirement for this technology. Logistics and budgetary challenges of obtaining fresh tissues notwithstanding, approaches to isolate cells from difficult or archived sources (e.g., interconnected neurons, flash frozen or cryopreserved tissues) using flow cytometry places the cells under stress, which substantially alters gene expression. One approach to overcome this limitation is to extract nuclei for RNA-seq thereby overcoming the difficulties involved in obtaining intact whole cells. MedGenome has developed a streamlined protocol to isolate nuclei from different sources of cells (fresh and cryopreserved cells). Here, we demonstrate that isolated nuclei can be used for downstream applications including gene expression profiling and epigenome profiling. Downstream analysis can be performed by modifying existing software solutions to account for the presence of pre-mRNA transcripts and obtain high-quality single nuclei transcriptome data. This method will also enable the exploration of biological features unique to nuclei, e.g., enrichment of certain transcripts and precursors of some noncoding RNAs.

Repurposing HiSeq 2500 sequencing systems as versatile fluidics and imaging platforms with PySeq2500

Technology Development

POSTER 576

Kunal Pandit[1,5], **Joana Petrescu**[2,3,4,5], **Miguel Cuevas**[2,3], **William Stephenson**[1], **Peter Smibert**[1], **Hemali Phatnani**[2,3,4], **Silas Maniatis**[1]

1 Technology Innovation Lab, New York Genome Center, New York, NY
2 Department of Neurology, Columbia University Irving Medical Center, New York, NY
3 Center for Motor Neuron Biology and Diseases, Columbia University Irving Medical Center, New York, NY
4 Center for Genomics of Neurodegenerative Disease, New York Genome Center, New York, NY
5 These authors contributed equally to this work.

We present PySeq2500, an open source code base and flow cell design enabling conversion of Illumina HiSeq2500 sequencing instruments for use as a flexible platform for advanced molecular assays. The HiSeq 2500 system was a legacy workhorse sequencer, with over 2000 systems in use as of 2018. Today, HiSeq 2500 systems have been obsolesced by new sequencers and Illumina has announced they will stop supporting the instrument and supplying the reagents in early 2023. However, the sequencer is still capable of performing as a robust, productive, and versatile benchtop system that includes precise XYZ stages, temperature control, fast imaging, and integrated fluidic handling. Using PySeq2500, we demonstrate coordinated control of the stages, pumps, valves, optics, and cameras of the HiSeq instrument for automated execution of iterative indirect immunofluorescence imaging (4i) of human and mouse tissue sections. We developed custom flow cells from off the shelf components to mount tissue sections and to interface with the HiSeq fluidics. Only minor modifications to the HiSeq were made to accommodate the custom flow cell and minimize dead volume. Simple PySeq2500 protocols enable versatile experimental designs in large, customizable flow cells involving simultaneous 4-channel image acquisition, temperature control, reagent exchange, and stage positioning. Our automated 4i method uses multiple cycles of staining with off the shelf antibodies, imaging, and antibody elution to build highly multiplexed maps of cell types and pathological features across multiple mouse and postmortem human spinal cord sections in each flow cell. This work demonstrates how PySeq2500 enables non-specialists to automate state of the art fluidics coupled imaging methods on widely available HiSeq 2500 systems.

HiFiViral SARS-CoV-2: a long-read kitted solution for genome surveillance

Technology Development

POSTER 577

Ting Hon[1], **Joan Wilson**[1], **Ian McLaughlin**[1], **Zev Kronenberg**[1], **John Harting**[1], **Meredith Ashby**[1], **Trang Dahlen**[1], **Janet Zeigle**[1], **Sarah Kingan**[1]

1 Pacific Biosciences (PacBio), Menlo Park, CA

The COVID-19 pandemic continues to be a major global epidemiological challenge with the ongoing emergence of new strain lineages that are more contagious, more virulent, drug resistant and in some cases evade vaccine-induced immunity. In response, the HiFiViral SARS-CoV-2 kit (PacBio; Menlo Park, California) was developed as a scalable solution for the Sequel II and Sequel IIe systems. Unlike amplicon sequencing, the HiFiViral SARS-CoV-2 kit uses tiled probes, resulting in robust genome coverage across a broad range of RNA input quantities and in the presence of new variants. The use of long, accurate HiFi reads enables comprehensive variant detection, including single nucleotide variants, indels, and structural variants, as well as phasing of variants if multiple strains are present in samples. The fully kitted solution contains all reagents needed for viral enrichment and barcoding up to 384 samples. Flexible scaling allows the user to pool 24 - 384 samples into a SMRT bell library to be run on a single SMRT Cell 8M. For high throughput labs, up to 8 SMRT Cells may be loaded on an instrument with no subsequent touch points. SMRT Link analysis reports include variant calling, genome coverage, detection of samples with multiple strains, and a plate performance summary for assay prep QC. File outputs include consensus sequences ready for submission to surveillance databases, reads aligned to the Wuhan reference, and consensus sequences aligned to the reference for advanced users. Here we demonstrate performance across a broad range of sample Ct inputs in control samples and nasopharyngeal samples in samples batches up to 384. HiFiViral for SARS-CoV-2 is a cost effective, convenient, and accurate method for viral sequencing, well-suited for scalable surveillance of a rapidly evolving virus to inform public health decision making.

A novel RNA-based single-cell spatial phenotyping workflow for immune profiling of whole slide samples

Technology Development

POSTER **578**

Nadav Yayon[1,2], **Julia Kennedy-Darling**[3], **Lydia Hernandez**[3], **Will Agnew-Svoboda**[3], **Ming Yu**[3], **Suyog Pathare**[3], **Arturo Gamboa-Aldeco**[3], **Garry Nolan**[4], **John C. Marioni**[1,2,5], **Sarah A. Teichmann**[1,6]

1 Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom
2 European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, United Kingdom
3 Akoya Biosciences, Menlo Park, CA
4 Stanford University School of Medicine, Stanford, CA
5 Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, United Kingdom
6 Cavendish Laboratory, University of Cambridge, Cambridge, United Kingdom

The thymus is the main site of the production of naïve T cells where selection processes ensure immune self-tolerance on one hand and reactivity to foreign antigens on the other. A majority of the human proteome is expressed in the thymus to make sure that self-antigens are not recognized by the T cell repertoire. Hence, the process of T cell maturation within the thymus is complex, coordinated, and heavily depends on spatial interactions between T cells and local stromal cells. Recent work by the Teichmann lab (Park et al., Science, 2020) identified a new T cell state within the thymus, GNG4+ CD8αα+, which is localized to the cortico-medullary junction and may play a significant role in shaping the thymocyte maturation landscape via recruitment and activation of dendritic cells. Spatial localization of a large number of targets will be required to fully capture the interactions of this cell type with others, as well as the many other cell maturation trajectories within the thymus.
Akoya Biosciences' spatial platforms enable the measurement of dozens of parameters with spatial context, across whole slides, and at single-cell resolution. To date, detection has focused primarily on protein targets; however, detecting RNA species directly holds in some cases, orthogonal information to protein, especially in cases where the protein target is not possible to detect due to lack of specificity, low expression levels, subcellular expression patterns as well as protein secretion and diffusion. Here, we present an analysis of thymus tissues using Akoya's new Hi-plex RNA detection assay. This workflow is enabled by the CODEX fluidics instrument, an end-to-end solution that enables single-cell, whole-slide spatial phenotyping at unprecedented speed. The ability to assess many RNA targets simultaneously within the tissue context without perturbing the integrity of the tissue allows the localization and identification of key cell types involved in thymic biology. Of note, the assessment of unconventional T cells, activated dendritic cells and the spatial correlation of immune cells with stroma cells enables a deeper understanding of unanswered questions related to the trajectory of T cell differentiation. With this technology, key cell types and associated sub-compartments are measured directly within the tissue context. This dataset demonstrates the value of single-cell, whole slide spatial RNA readouts.

Piloting NGS-based newborn screening

Technology Development

POSTER **579**

M.R. Nelen[1], **D. Westra**[1], **M.B.G. Kiewit**[2], **R.J. Sinke**[2], **M.R. Heiner Fokkema**[3], **B. Raddats**[2], **E. Voorhoeve**[4], **F.J. van Spronsen**[5]

1 Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands
2 Department of Medical Genetics, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands
3 Department of Laboratory Medicine, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands
4 Centre for Health Protection, National Institute for Public Health and the Environment, Bilthoven, Netherlands
5 Department of Metabolic Diseases, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands

Newborn blood screening (NBS) at a population base is a public health program that aims to identify newborns in whom immediate treatment can save their life or improves outcome. The current Dutch NBS has shown its effectiveness for 18 diseases including some largely non-genetic diseases, e.g. congenital hypothyroidism, and 13 inborn errors of metabolism (IEM) that are all genetically based. Currently, this biochemical-first strategy is largely using tandem mass spectrometry (MS/MS) analyzing dried blood spots (DBS) for abnormal metabolite concentrations in blood on filter paper that are collected between 3 and 7 days after birth. Notwithstanding the success, this approach also results in a relatively high false positive rate and false negative results. Even more important, many other IEM cannot be added to the NBS program due to the lack of biochemical footprints in DBS. Using next generation sequencing (NGS) techniques as start in NBS (NGS-first) can overcome the lack of a biochemical footprint of these IEM in DBS, may decrease the number of both false positive and false negative results.The present pilot study focusses on such technical aspects.

The overall aim elaborates on the following overall task statements of the three objectives:

1. Select IEM not only theoretically appropriate to be screened using rapid NGS-first analysis (either using WES or WGS), but also with the knowledge that the selected genes will have an analytical and bio-informatic workflow necessary in NGS-first NBS and decide on the acceptable level of sensitivity and specificity of the NGS-first based NBS of the selected genes (WP1).
2. Develop the technical procedures to a NGS-first based NBS of the in WP1 selected genes of IEM to the extent that it is technically and analytically ready for a development trajectory in the NBS reference laboratory of the RIVM (WP2).
3. Investigate the necessary steps to be taken in a NGS-first based NBS trajectory in which infants with pathological mutations within the genes of the in WP1 selected IEM -and of which the methodology have been developed in WP2- can be flagged by the NBS reference laboratory of the RIVM (targeted WES or WGS) and can be referred to the metabolic clinic without delay, while the DBS of infants with genotypes that cannot (yet) be fully considered pathological are investigated by biochemistry as second tier in the NBS program (WP3).

Introduction to CRISPRclean™: the in vitro use of CRISPR-CAS to improve NGS applications

Technology Development

POSTER 580

Azeem Siddique, **Keith Brown, Gaia Suckow, Dhanya Rhamachadran, Jon Armstrong**

Jumpcode Genomics, San Diego, CA

CRISPRclean™ technology uses the sequence specific endonuclease activity of the CRISPR CAS system to inhibit amplification and sequencing of abundant molecules in NGS libraries. Applications range from host depletion for increased sensitivity of pathogen detection, removal of abundant transcripts from single cell, spatial and ISO-Seq applications, removal of repetitive sequences from complex genomes to reduce cost and increase coverage over informative genetic elements and can also be used to deplete allele specific sequences for methylation analysis or wild type removal from liquid biopsy assays. One major benefit of the workflow is the incorporation of CRISPRclean™ post library generation. This enables a dramatically reduced sample input into standard library construction techniques and the ability to serially deplete abundant sequences for further exploration of low abundant molecules in a sample. Because the platform uses single guide RNAs of about 100-120 bp in length, these can be produced at massive scale and breadth using commercially available silicon based DNA writing techniques and In vitro transcription. In this presentation we explore both the lower boundaries of sample inputs at single cell levels (10pg), including the lack of perturbations of non-depleted molecules through ERCC spike in controls and log^2 transcription profiles across a wide range of inputs in whole transcriptome sequencing. In addition, we explore the current upper boundaries (or lack of an upper boundary) for guide complexity by presenting data on guides sets greater than 500,000 unique sgRNAs per reaction in applications of rare disease diagnostics and high throughput genotyping in complex plant genomes.

An integrated single-cell multiomics approach to characterize mRNA, intracellular proteins, and surface proteins using intracellular AbSeq and the predesigned, lyophilized, BD® AbSeq Immune Discovery Panel

Technology Development

POSTER 581

Hye-Won Song[1], **Samatha Vadrevu**[1], **Jody Martin**[1], **Eileen Snowden**[2], **Timothy Zheng Zhu**[3], **Adam Wright**[1], **Margaret Nakamoto**[3], **Shannon Dillmore**[2], **Aruna Ayer**[3]

1 BD Biosciences, San Diego, CA
2 BDTI, Durham, NC
3 BD Biosciences, San Jose, CA

High-throughput single cell RNA sequencing (scRNA-seq) has recently emerged as a powerful tool for profiling complex cell populations. While traditional scRNA-seq captures information at the transcript level, recent technology using DNA-barcoded antibodies enable multiomic analysis, examining protein alongside mRNA. Although this approach enables analysis of a large number of protein markers, leading methods are currently limited to the study of surface proteins. Understanding signal-transduction and transcriptional pathways requires the ability to analyze intracellular proteins because intracellular protein expression, post-transcriptional, and post-translational modifications often govern these processes. The ability to simultaneously analyze intracellular proteins alongside the transcriptome is complicated by the cell fixation and permeabilization processes necessary to access intracellular epitopes, which can negatively impact RNA assay sensitivity. Furthermore, nonspecific binding of antibody-conjugated oligos to intracellular nucleic acids can generate high background signal that masks true epitope-antibody interactions.

To address these issues, here we generated a protocol that can recover high quality RNA alongside specific intracellular and surface protein information from single cells. Using this protocol in conjunction with the BD Rhapsody System we simultaneously profiled 6 intracellular proteins in stimulated human immune cells along with whole transcriptome analysis and 30 surface proteins at the single cell level. In this proof-of-concept study, we identified predicted patterns among these intracellular markers, including increased phosphorylation of AKT protein correlated with the proliferation marker Ki67 upon immune stimulation, and inverse correlation of AKT phosphorylation with cleaved PARP expression, an apoptotic marker. These results are concordant with the role of AKT activation in cell survival and proliferation signaling. Together these results show that analysis of intracellular proteins together with surface proteins and the transcriptome is possible using single-cell sequencing technologies and can enable a deeper understanding of cellular states.

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

Multimodal cell segmentation for spatial in-situ imaging of tissues

Technology Development

POSTER 584

Lidan Wu[1], **Aster Wardhani**[1], **Patrick Danaher**[1], **Youngmi Kim**[1], **Shanshan He**[1], **Zachary Lewis**[1], **Carl Brown**[1], **Rustem Khafizov**[1], **Dwayne Dunaway**[1], **Michael Rhodes**[1], **Joseph Beechem**[1]

1 NanoString Technologies, Seattle, WA

Spatial transcriptomics technologies, which produce single-cell gene expression data paired with cell locations, have opened a new frontier in biology. Accurate cell segmentation that assigns transcripts to cell locations is critical to data quality, but very challenging for tissue sections where cells are tightly packaged with shared, 3D boundaries and uneven morphology staining. To address this we have developed a multimodal cell segmentation pipeline that automatically does image preprocessing, machine-learning-augmented cell segmentation and transcript-based segmentation refinement. We demonstrate our pipeline on spatial transcriptomics datasets from various FFPE tissue sections.

Firstly, our pipeline takes tissue images stained with both nuclear and membrane markers (DAPI, CD298/PanCK/CD3) to perform rescaling, normalization and boundary enhancement. The preprocessed images are fed into pre-trained Cellpose neural network models for both nuclear segmentation using nuclear channel only and cytoplasm segmentation using combined nuclear and membrane channels. Outputs of the two segmentation methods are combined by analyzing the intersection-over-union scores to: (a) mitigate issue of non-uniform staining in either channel; and (b) enable subcellular compartment analysis of the spatial transcriptomics data.

Secondly, our pipeline exploits the spatial profile of transcripts on the same tissue section for rapid segmentation refinement. The image-based cell segmentation enables single-cell typing/clustering of transcript profiles. Individual transcripts are then scored for goodness-of-fit within their respective cells, based on the probability of each gene belonging to each cell type and the spatial dependency of transcript score. As confirmed by the membrane-stained images, cells with boundary errors at the junction of different cell types, exhibit strong spatial dependency in their transcript score profile and thus can be easily identified. Our pipeline further identifies the spatially connected groups of transcripts with low goodness-of-fit within incorrectly segmented cells. A set of heuristic rules on neighborhood cell typing and transcript number are then applied to the identified transcript groups to decide on the re-segmentation actions, like merging, splitting and trimming. The re-segmented results show no significant spatial dependency on transcript score of individual cells, suggesting the successful correction of poorly segmented cells.

Cell segmentation is critical for data quality in spatial transcriptomics. The pipeline presented here harnesses information from both morphology images and mRNA locations to generate an automated and robust cell segmentation algorithm across different tissue types. The technique is computationally tractable with > 1 million cells, making it viable for even the biggest spatial transcriptomics experiments.

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

Automated processing of solid tissues into single cells or nuclei with the Singulator™ 100 system: scRNA-Seq, snRNA-Seq, and ATAC-Seq data on human and mouse tissues

Technology Development

POSTER 585

John Bashkin, **Kaiwan Chear, Stevan Jovanovich , Nabiha Khan, Bruce Leisz, Danielle Meyer, Nathan Pereira**

S2 Genomics, Livermore, CA

Single-cell sequencing is revealing the next level of complexity of biological systems. The ability to measure the heterogeneity of single-cell gene expression and chromatin structure cell by cell has led to the discovery of new cell types and expanded our understanding of the diverse regulatory pathways and states of individual cells. Single-cell sequencing workflows for multiple applications require reproducible generation of high-quality single-cell or nuclei suspensions.

S2 Genomics has developed and commercialized the patented Singulator™ 100 System to automate the dissociation of solid tissues into single cell or nuclei suspensions. The Singulator disrupts solid tissues in single-use cartridges using coupled enzymatic/chemical dissociation and mechanical disruption.

Data will be presented showing the use of the Singulator 100 to prepare samples for subsequent analysis by scRNA-Seq, snRNA-Seq, and ATAC-Seq on multiple human and mouse tissue samples. The impacts of sample processing conditions on RNA and sequencing library quality will be discussed, along with data on cell type representation and transcriptional differences observed between cells and nuclei isolated from the same samples.

Supported in part by National Institutes of Health award number R41 HG010129-01 to Dr. Jovanovich.

Streamlining NGS-based assays for genome engineering

Technology Development

POSTER 586

Jack T. Leonard[1], **Kathleen Collins**[2]

1 seqWell, Inc., Beverly, MA
2 University of California, Berkeley, Berkeley, CA

Genome engineering using programmable fusions of DNA binding and nuclease domains (PDNs) to insert transgenes of choice has enabled research and holds promise as a general approach for gene therapy. However, the specificity of target site use can vary. For example, genome editing by TAL Effector and Zinc finger-based PDNs is influenced by chromatin state, and genome editing by CRISPR/Cas9 can be accompanied by off-target mutagenesis. Repair of Cas9-induced double-stranded DNA breaks at sites with near-matching target sequence occurs via nonhomologous end-joining, producing unwanted off-target mutations. Furthermore, DNA introduced to cells can insert at unpredictable locations, obliging a method to efficiently screen for all genome insertion sites in an unbiased manner. UDiTaS (Giannoukos et al., 2018) and GUIDE-Seq (Tsai et al., 2015) belong to a family of library prep methods aimed at detecting and quantifying these types of on- and off-target effects of CRISPR/Cas9 and other genome engineering enzymes.

In general, it is effort-intensive, slow, and bias-prone to generate libraries of genomic DNA by fragmentation, size selection, end-repair and ligation prior to PCR. Library amplification is often followed by inefficient hybridization-based enrichment of sequences of interest to reduce sequencing cost. seqWell UMI-tagging reagents circumvent these limitations by enabling streamlined, highly multiplexed, transposase-based workflows for putative target site or unbiased insertion site profiling. In the seqWell workflow, an initial tagging step is performed to add a full-length adapter carrying P5, an i5 index, a unique molecular identifier (UMI) upstream of the transposase-binding region and an Illumina Nextera®-style Read1 sequencing primer. Next, a single or multiplexed set of user-customized PCR primers is used in combination with P5 primer to amplify genomic DNA between the custom primer(s) and the nearest transposition site. By titrating the ratio of UMI-tagging reagent to input genomic DNA, target or insertion site regions of different length ranges can be obtained. The UMIs carried on the full-length tagging adapters allow for correction of PCR bias introduced during library amplification and thereby provide an unparalleled comprehensive, quantitative read-out of both location and relative number of on-target and off-target sites in complex populations of cells. We demonstrate application of the seqWell UMI-tagging reagents and workflow to characterize and quantify on-target versus off-target transgene insertions into the human genome.

A hybridization capture-based approach for novel gene fusion discovery

Technology Development

POSTER 587

Katelyn Larkin, **Jessica Sheu, Kevin Lai, Shale Dames, Ashley Dvorak**

Integrated DNA Technologies, Coralville, IA

RNA-Seq allows researchers to quantify gene expression and identify sequence alterations such as isoforms, alternative splicing events, and gene fusions across different tissues or conditions, which in turn provides a more complete understanding of transcriptome dynamics during diseased and normal states. One of the challenges with RNA-seq is the sequencing cost associated with the high coverage needs of low-expressed transcripts or rare structural variations. By using a targeted sequencing approach, sequencing reads can be focused on specified regions of interest in the transcriptome, allowing for deeper sequencing of these areas while omitting undesired regions and other highly expressed or dominant transcripts such as certain housekeeping genes and ribosomal RNA (rRNA). Gene fusions are important biomarkers for researchers as they are known to be cancer drivers; structural rearrangements like translocations and deletions can lead to the formation of hybrid genes, or gene fusions, while non-structural rearrangements such as transcript read-through of mRNA splicing can also lead to the formation of gene fusions. The capability for deeper sequencing with lower cost enables researchers to use targeted RNA-seq as a tool for the identification of rare fusion events and discovery of novel fusion transcripts that may otherwise be too costly and difficult to identify.

Here we describe a hybridization capture-based targeted RNA-Seq approach that can be used for both the study of known gene fusions and the discovery of novel gene fusions while still allowing quantitative assessment of gene expression. We show how an enhanced approach to a hybridization capture probe design allows researchers to identify a similar number of novel fusion-supporting sequencing reads as would be seen when designing capture probes specifically for known gene fusions.

Disclaimer:
For research use only. Not intended for use in diagnostic procedures. Unless otherwise agreed to in writing, IDT does not intend these products to be used in clinical applications and does not warrant their fitness or suitability for any clinical diagnostic use. Purchaser is solely responsible for all decisions regarding the use of these products and any associated regulatory or legal obligations.

Spatial mapping of the transcriptome of peripheral blood mononuclear cells obtained from liquid biopsy

Technology Development

POSTER 588

Jagadish Sankaran[1], **Giovani Claresta Wijaya**[1], **Shelly Hsiao-Ying Cheng**[2], **Nirmala Arul Rayan**[1], **Iain Bee Huat Tan**[1], **Kok Hao Chen**[1], **Shyam Prabhakar**[1]

1 Genome Institute of Singapore, Nucleos, Singapore
2 Veranome Biosystems, Nucleos, Singapore

Liquid biopsies are minimally invasive and low risk clinical procedures to collect blood for basic science and diagnostic investigations. The cost, labor and time efficiency of liquid biopsies enable the monitoring of disease progression or response to treatment at regular intervals. A typical workflow in liquid biopsy involves collection of blood, and then lysing to quantify biomarkers leading to a loss of positional biomarker information. We overcome this limitation by using multiplexed Fluorescence In Situ Hybridization (multiplexed FISH) wherein spatial gene expression is quantified at sub-cellular resolution. Cyclic FISH assays offer novel opportunities to identify disease states based on not only abundance but also based on sub-cellular mislocalization of transcripts.

We developed a workflow to perform multiplexed FISH on peripheral blood mononuclear cells (PBMCs), and used it to map the spatial expression of 140 immune markers. Image segmentation yielded cellular-resolution spatial transcriptomes of the major blood cell populations: T cells, B cells, Natural Killer cells and monocytes. Having validated the fidelity of cellular transcriptome measurement, we proceeded to investigate sub-cellular localization of transcripts. In most blood cells, the nucleus occupied a large proportion of the volume, and thus transcripts were concentrated in a shallow cytoplasmic layer close to the cell membrane. We are currently multiplexing the assay to quantify the ratio of nuclear to cytoplasmic abundance (RNCA) of various transcripts in the transcriptome. We hypothesize that RNCA will serve as a metric to identify disease stage and disease state, and thus non-invasively monitor response to treatment.

Deep learning morphology analysis identifies and enriches carcinoma cells from body fluids in real-time for cytological and molecular analysis

Technology Development

POSTER 589

Anastasia Mavropoulos[1], **Jeanette Mei**[1], **Janifer Cruz**[1], **Kiran Saini**[1], **Nianzhen Li**[1], **Chassidy Johnson**[1], **Thomas J. Musci**[1], **Jianyu Rao**[2], **Mahyar Salek**[1], **Maddison (Mahdokht) Masaeli**[1]

1 Deepcell Inc., Menlo Park, CA
2 University of California, Los Angeles (UCLA), Los Angeles, CA

Detection and characterization of cancer cells in body fluid specimens is of great importance in personalized medicine but is hindered by low sensitivity and specificity of detection in a high background of non-malignant cells. We addressed this limitation by applying our novel platform to detect and enrich carcinoma cells from body fluids. The platform combines high-resolution imaging of single cells in flow with real-time morphology based cell identification and sorting yielding label-free, intact cells using deep learning models, which we call Computational Sorting and Mapping of Single cells (COSMOS). We have trained deep convolutional neural network (ConvNet) models on millions of single cell bright field images to identify and discriminate between various malignant, stromal and immune cell subtypes. The platform output yields an enriched population of label-free cells that are compatible with molecular analysis with improved sensitivity compared to a pre-enriched sample.

To generate models specific to ascites and pleural effusions we used >3 million training cell images from malignant body fluid samples (cancer, mesothelial, histiocytes, and inflammatory cells). Using our platform and model to analyze body fluid samples (ascites, pleural effusions) from patients with a history of malignancy we detected carcinoma cells in 100% (n=18) of carcinoma positive cytology samples (certified cytologist readout was ground truth). Flow cytometry was used to verify model performance by comparing the percent carcinoma cells based on EpCAM/Claudin-4 profile to percent malignant cells determined by the model. 14/18 (77.8%) of carcinoma positive cytology samples had comparable (within 10%) or higher percent carcinoma cells detected using the model compared to FACS.

We then went on to demonstrate the capacity for sorting and enrichment of malignant cells on the platform using SNP and CNV analysis. First a breast cancer cell line was spiked into non-cancer cells at 1% and we demonstrated a 60-fold enrichment using SNP analysis on both pre- and post-enriched samples. Similarly with a patient ascitic fluid sample we demonstrated an approximate 4-fold enrichment. The results confirm enrichment of carcinoma cells based on model classification and real-time sorting and we have shown increased sensitivity of CNV analysis.

This work demonstrates that our platform (COSMOS) can enhance current cytology workflows by achieving accurate classification and enrichment of label-free malignant cells that can be further analyzed by molecular analysis which could greatly improve precision medicine for cancer patients.

Automation, acceleration, and parallelization of nuclei isolation from tissue for subsequent single-nuclei sequencing

Technology Development

POSTER 590

Jana Soyka[1], **Carsten Poggel**[1], **Lea Eich**[1], **Lilian Andrea Martinez Carrera**[1], **Nojan Jelveh**[1], **Ariadna Pascual Velazquez**[1], **Erin Christensen**[1], **Carina Emery**[1], **Chintda Santiskulvong**[2], **Xue-Ying Song**[2], **Alex Rajewski**[2], **Olaf Hardt**[1]

1 Miltenyi Biotec B.V. & Co. KG, Bergisch Gladbach, Germany
2 Cedars-Sinai, Applied Genomics, Computation, and Translational Core, Los Angeles, CA

Single-cell sequencing applications have accelerated breakthroughs in life sciences and transformed our understanding of biology, leading to significant impacts for disease treatments. However, many sample types are not suitable for dissociation into a viable single-cell suspension, which is necessary for whole transcriptome single-cell RNA-sequencing (scRNA-seq). Unsuitable sample types include snap-frozen tissue stored in biorepositories and tissue with large, fragile cell types, such as adult neural tissue. To investigate these challenging sample types, many researchers have turned to single-nuclei RNA-sequencing (snRNA-Seq), a substitute approach for scRNA-Seq widely used in other applications. Current protocols for isolating nuclei from complex tissues typically include labor-intensive, manual methods for mechanical disaggregation of the tissue and debris removal, and are not scalable to run multiple samples in parallel. Here, we demonstrate an end-to-end workflow solution, including nuclei extraction and isolation for reproducible snRNA-Seq analysis using frozen archival human tumor tissue and adult mouse brain. Our workflow starts with automated nuclei extraction of up to eight samples in parallel on the gentleMACS™ Octo Dissociator. Nuclei suspensions are then cleaned of debris using two methods, either magnetic MACS® MicroBeads or sorting on the MACSQuant® Tyto® Cell Sorter. Nuclei quality was checked before and after sorting by microscopy and flow cytometry. Finally, to evaluate the effects of debris removal on data quality and to show that nuclei isolation and sorting indeed improve the library complexity and purity, we used the 10x Genomics platform to generate 3' gene expression libraries out of nuclei suspensions obtained from before and after isolation or sorting, respectively. In conclusion, our experiments demonstrate in several approaches that nuclei isolation workflows can be improved by automating nuclei extraction for scalability and reproducibility and MACS MicroBead isolation or flow sorting to clean nuclei suspensions of cellular debris.

Rapid identification of DNA fragments through direct sequencing with electro-optical zero-mode waveguides

Technology Development

POSTER 591

Fatemeh Farhangdoust[1], **Mohammad Amin Alibakhshi**[2], **Feng Cheng**[3], **Wentao Liang**[2], **Yongmin Liu**[3], **Meni Wanunu**[1,2]

1 Department of Bioengineering, Northeastern University, Boston, MA
2 Department of Physics, Northeastern University, Boston, MA
3 Department of Electrical and Computer Engineering, Northeastern University, Boston, MA

In contrast to sequence-specific techniques such as polymerase chain reaction (PCR), DNA sequencing does not require prior knowledge of the sample for surveying DNA. However, current sequencing technologies demand high inputs for a suitable library preparation, which typically necessitates DNA amplification, even for single-molecule sequencing methods. Here, we present electro-optical zero-mode waveguides (eZMWs), which can load DNA into the confinement of zero-mode waveguides with high efficiency and negligible DNA fragment length bias. Using eZMWs we observe highly efficient voltage-induced loading of DNA fragments of various sizes from ultralow inputs (ng-to-pg levels). We demonstrate rapid DNA fragment identification by burst sequencing of short and long DNA molecules (260 bp and 20,000 bp) loaded from an equimolar pM-level concentration mixture in just a few-minutes. Our device allows further studies in which low-input DNA capture is essential, for example, in epigenetics, where native DNA is required for obtaining modified base information.

A highly reproducible and robust system for subcellular in-situ single-cell spatial transcriptome profiling of human brain tissue

Technology Development

POSTER 592

Shauna Edridge[1], **Chloe Kim**[1], **Cindy Chen**[1], **Eric A. Lee**[1], **Ustun Tulu**[1], **Chang Choi**[1], **Eric P. Lee**[1], **Yun-Ching Chang**[1], **Debjit Ray**[1], **Marc Glazer**[1]

1 Applied Materials, Santa Clara, CA

The human brain is a highly complex organ with a multitude of cell types working synchronously. Without understanding the spatial transcriptomic profiling at a single cell level, it is impossible to fully understand how brain cells function. In this study, we demonstrate the application of a fully integrated, end-to-end solution for in-situ single-cell transcriptome profiling with human brain tissue. Utilizing multiplexed FISH hybridization-based chemistry, Veranome's VSA-1 system capable of subcellular resolution imaging, and proprietary cell profiling software, we demonstrate highly robust and reproducible results with human brain tissues, including samples that contain substantial autofluorescent lipofuscin aggregates. Currently, the Veranome workflow is highly multiplexing and able to provide informative results for more than 500 genes in tissues. In this study, approximately 200 genes were targeted. This method does not rely on the use of tissue clearing, tissue expansion, or chemical autofluorescence quenching techniques for reduction of lipofuscin autofluorescence. We identified several neuro cell type specific gene expression patterns which correlate with published scRNA-seq and in-situ hybridization data. Apart from spatial localization images, the cell profiling software is also able to qualify, analyze, and visualize spatial transcriptomic data. For example, comprehensive meta-analysis can be accomplished on cell types such as astrocytes, endothelial cells, microglia, neurons, oligodendrocytes as well as other cell and sub-cell types. We generated results that show that the workflow can be combined with multiplexed immunofluorescence staining and that the specificity of cell type RNA expression data spatially coincides with cell type immunostaining. We have demonstrated that anti-GFAP immunostaining of astrocytes can be visualized in combination with RNA expression data for GFAP, AQP4 and other astrocyte expression markers while neurons can be profiled with anti-NeuN immunostaining in combination neuronal expression markers such as RELN. The streamlined, robust and highly reproducible VeranomeTM workflow can facilitate in-situ single-cell transcriptome profiling with human brain tissue and many other applications.

In situ single-cell transcriptomic imaging in formalin-fixed paraffin-embedded (FFPE) tissues with MERSCOPE

Technology Development

POSTER 593

Jiang He[1], **George Emanuel**[1], **Justin He**[1]

1 Vizgen, Cambridge, MA

Formalin-fixed paraffin-embedded (FFPE) tissues are the most widely used clinical sample types in histology and molecular diagnosis. However, due to RNA degradation and protein crosslinking, FFPE samples are often not compatible with single-cell transcriptomic analysis. To accurately profile the gene expression in FFPE samples in situ, a spatial transcriptomics technique with high detection efficiency and single molecule resolution is required. Vizgen's in situ genomics platform MERSCOPE-TM, built on multiplexed error robust in situ hybridization (MERFISH) technology, enables the direct profiling of the spatial organization of intact tissue with subcellular resolution. Here, we present a workflow to perform in situ single-cell transcriptomic imaging in FFPE tissue sections, and demonstrate MERFISH imaging in various FFPE samples, including mouse brain, liver, kidney, lung, and human cancer. Specifically, using a panel of 483 genes that includes canonical cell type markers and G protein-coupled receptors (GPCRs) to spatially profile gene expression across the mouse brain, we demonstrated that MERSCOPE was able to create a molecular and cellular atlas of the FFPE mouse brain, with no drop of the detection efficiency as compared to fresh frozen sample types. Notably, many lowly expressed non-sensory GPCRs such as Oxtr, Tshr, and Insr, were successfully located in the FFPE mouse brain. We also exhibited a strong correlation (r~=0.9) between the number of transcripts detected to bulk RNA sequencing. By clustering cells based on their gene expression, we identified different cell types and mapped their spatial organization. The cell types identified in FFPE mouse brain have excellent correlation with the cell types identified in fresh frozen samples. Furthermore, using an oncology-focused gene panel, we demonstrated that MERSCOPE was able to accurately profile gene expression in situ and map cell types in archival clinical human samples. We were able to map and catalog different immune cell types within the tumor microenvironment and systematically characterize the gene expression among cells. Spatially resolved transcriptomic profiling of FFPE samples at single-cell level provide enormous opportunities in a wide range of biomedical research areas and will likely open up many applications to study human diseases.

A new spin on sample preparation: rapid, scalable isolation of nuclei for single-cell genomics

Technology Development

POSTER 594

Michael Gibbons[1], **Alaina Puleo**[1], **Jill L. Herschleb**[1], **Sarah E.B. Taylor**[1]

1 10x Genomics, Pleasanton, CA

Sample preparation remains a significant roadblock to implementation and scaling of single cell technologies in clinical and translational research. Time consuming, custom dissociation protocols performed on fresh tissue by experts in a particular disease are often utilized to obtain high quality single cell data, which leads to low sample throughput and high variability driven by logistical challenges and operator effects. Additionally, many samples that are of interest for clinical and translational studies arrive in research labs snap frozen, leaving nuclei extraction as the only viable path to generate single cell/nuclei data. To overcome these challenges, we present a scalable, easy to use, spin column-based nuclei isolation platform that generates single nuclei that are ready for input into downstream 10x Genomics single cell assays in under an hour.

To highlight the utility of this platform we processed eight unique human tumors: skin (melanoma), breast, ovary, prostate, lung, kidney, colorectal, and pancreas. These samples were processed in parallel, going from frozen tissue to single nuclei suspensions loaded into the 10x Chromium Controller within 90 minutes. The resulting single nuclei were run through the 10x Genomics Single Cell Gene Expression, Single Cell Immune Profiling, and Single Cell ATAC workflows. Data from all single cell gene expression analyses retained high complexity in terms of UMI and genes per cell across all tumor types tested. Expected cell type clusters, including tissue specific tumor cells, surrounding/infiltrating immune cells and stromal cells were easily identified and profiled via gene expression analysis, with even fragile cell types often missing in single cell data (e.g. mature adipocytes) preserved and profiled. Additionally, epigenetic changes as measured by Single Cell ATAC profiling were consistent with both general and tissue-specific epigenetics of each cancer type analyzed.

In summary, this platform streamlines sample preparation for single cell studies from snap frozen tissue, creating opportunities for these samples to be used much more routinely. The findings presented here demonstrate that this fast and easy to use nuclei isolation system provides a repeatable platform for generating high quality single nuclei data at scale for clinical, translational and basic research studies alike.

A novel RNA-based single-cell spatial phenotyping solution for immune-profiling in cancer

Technology Development

POSTER 595

Alex Tsankov[1], **Julia Kennedy-Darling**[2], **Lydia Hernandez**[2], **Will Agnew-Svoboda**[2], **Ming Yu**[2], **Suyog Pathare**[2], **Arturo Gamboa-Aldeco**[2], **Garry P. Nolan**[3], **Samouil Farhi**[4]

1 Icahn School of Medicine at Mount Sinai, New York, NY
2 Akoya Biosciences, Menlo Park, CA
3 Stanford University School of Medicine, Stanford, CA
4 Broad Institute of MIT and Harvard, Cambridge, MA

The immune state of the tumor microenvironment (TME) can dictate whether a patient is responsive to immunotherapy. Multiple factors, including T cell infiltration, myeloid cell population and pH, all contribute to this propensity. For non-small cell lung cancer (NSCLC) patients, TME associated T regulatory cells have been implicated in immune escape—it was recently shown that patients with low density CD8+Foxp3+ infiltrates had higher overall survival rates. This parameter alone, while correlative, does not explain every case and highlights the need to measure both spatial localization and multiple cell types to decipher the underlying immune state of the tumor.

Highly multiplexed spatial technologies have emerged to address this need. Here we demonstrate the application of a new Hi-Plex RNA spatial solution from Akoya Biosciences for spatial phenotyping of immune cells in NSCLC patients. This targeted approach uses a hybridization-based methodology to reveal subsets of markers across multiple cycles of fluidics and high-speed imaging. As a hybridization based method, this approach detects individual RNA molecules, achieving single-cell resolution. Cell type annotation based on the detected markers can be performed, and, since sample integrity is maintained through the experiment, cell type arrangements—or neighborhoods—may be evaluated. In this demonstration, markers were selected based on relevant genes implicated in the TME of NSCLC patients. Immune, tumor and stromal regions of the tissue were assessed to map the neighborhoods present within the TME. This technology could advance understanding of the underlying tumor progression, immune escape, and propensity of patients to respond to different therapies.

A novel RNA-seq workflow for comprehensive transcriptome profiling

Technology Development

POSTER 596

Jason Ma[1], **Cindy Wan**[1], **Mark Shulewitz**[1], **Xiangdong Meng**[1], **Xiao-Song Gong**[1], **Steven Okino**[1], **Stephen Bolaris**[2], **Jeff Bowie**[2], **Zach Burkett**[2], **Yan Wang**[1], **Cissy Jiang**[1*]

1 Bio-Rad Laboratories, Life Science Group, Hercules, CA
2 Bio-Rad Laboratories, Digital Biology Group, Pleasanton, CA
*Presenting author

RNA-Seq is a powerful technique to measure the presence and abundance of RNAs in a cellular sample. We present a unique and innovative 3-step RNA-Seq workflow that combines (1) streamlined RNA-Seq library preparation; (2) multiplex compatible ribodepletion, and (3) an integrated web-based analysis pipeline for sequencing data processing. The RNA-Seq library preparation can be achieved in ~3 hours (including amplification step). The streamlined workflow is enabled by using a novel reverse transcriptase in combination with a unique ligation-free adaptor addition chemistry. The resulting cDNA library can be processed using a post-library ribodepletion kit, which depletes cDNA originating from ribosomal RNA from up to 96 indexed cDNA libraries in one tube with high efficiency and specificity. The entire workflow from RNA to indexed ribodepleted cDNA library can be completed in ~5 hours (including amplification step). These libraries are compatible with all Illumina sequencers for paired-end sequencing. By using a unique molecular identifier (UMI) in the adaptor, we demonstrate that the PCR duplication rate observed with these libraries is significantly lower than other on-market RNA-Seq products, especially when the input RNA quantity is low. The RNA-Seq results are reproducible, quantitative and can be analyzed using an integrated informatics solution. Data can be further analyzed using other downstream data analysis tools, such as differential gene expression, to reveal biologically relevant differences between RNA samples. We further show that this novel workflow can accurately identify RNAs that are differentially regulated during cellular differentiation. This integrated and innovative RNA-seq workflow provides a powerful and efficient option that enables transcriptome profiling to become a more standard and mainstream approach.

CRISPRclean™ powered host depletion improves pathogen detection in a shotgun RNA-Sequencing–based universal infectious disease test

Technology Development

POSTER 597

Keith Brown, **Azeem Siddique**, **Jon Armstrong**, **Yvain Desplait**

Jumpcode Genomics, San Diego, CA

It took 42 days from the first reported clinical case of COVID-19 until the viral genome sequence was provided to the public. This critical information release acts as the starting point for diagnostic test development and mRNA vaccine generation. The first EUA approved RT-PCR tests in the United States were made available after 70 days on February 8th 2020, and came with a contamination failure. Shotgun metagenomic NGS sequencing is the only technology capable of unbiased detection from previously unknown pathogens available on day ZERO (0) for future pandemics. Beyond detection, meta-genomic sequencing enables full genome sequence assembly, the ability to catalog co-infections, and provides valuable data on host response through host transcriptional profiling. Sensitivity of shotgun mNGS is hindered by abundant eukaryotic (host) and prokaryotic rRNA sequences from nasal swabs. Using a guide set of over 12,000 sgRNAs targeting abundant human and bacterial ribosomal RNA sequences in the CRISPRclean™ workflow, we show that sequence specific cleavage results in read space shifting from ribosomal "contaminants" to provide an average 10-fold increase in sensitivity to detect non-human sequences. In this study, 50 COVID positive samples and 10 negative samples by RT-PCR, from a clinical testing lab in California, were subjected to amplicon based sequencing and shotgun metagenomic sequencing with and without rRNA depletion using CRISPRclean™. We report detection sensitivity comparisons to RT-PCR, variant concordance to amplicon sequencing and an early look at host response through transcriptional profiling across technical replicates. In addition to HCoV2, this study has identified a number of co-infections spanning NIAID taxa flagged as priority levels A, B, and C. Data, protocols, and analysis pipelines will be provided as supplemental information accessible from a link on the poster. The protocol, including a strand specific library prep with Unique Dual Indexes has been fully automated from extraction through library construction and depletion on FDA 510K approved liquid handlers. The full solution represents a viable approach for pandemic preparedness by enabling immediate detection of all pathogens at the site of an outbreak or points of entry into or out of the country. Bridging the time gap from initial outbreak to the availability of low cost, specific pathogen detection at scale could save lives in the future.

Spatial analysis of formalin-fixed, paraffin-embedded (FFPE) samples using Veranome'sTM workflow

Technology Development

POSTER **598**

Erica Teo[1], **Chang Choi**[1], **Ustun Tulu**[1], **Chloe Kim**[1], **Quentin Klueter**[1], **Zheng Guang Ang**[1], **Li Zhen**[1], **Marc Glazer**[1]

1 Veranome Biosystems LLC, an Applied Materials company, Mountain View, CA and Nucleos, Singapore

Veranome's end-to-end omics solution has been proven to generate high-quality and reproducible data using cryosections in cells, organoids, and tissues. In this study, the capability of Veranome's solution is extended to formalin-fixed, paraffin-embedded (FFPE) samples. FFPE samples are widely used in clinical diagnostics and research. The process of fixation, embedding and storage for extended time periods is known to negatively impact RNA quality of these samples. We overcame these challenges and demonstrated highly reliable and reproducible results with mouse and human brain tissues. Veranome's multiplexed FISH hybridization-based workflow seamlessly integrates with the VSA-1 system, capable of subcellular resolution imaging, Veranome's reagents products and proprietary software. Currently, Veranome's cyclic FISH workflow is highly multiplexing and able to provide informative results to over 500 genes in tissues with the capability to go higher as needed. Additionally, this technique does not rely on the usage of tissue clearing nor tissue expansion methods for removal of autofluorescence, which was especially evident in the human brain tissue samples. Due to the cross-links introduced during the formalin fixation steps, antigen retrieval was conducted to ensure probes could penetrate the samples. We demonstrated that Veranome's workflow was compatible with both enzymatic digestion and heat-induced epitope removal methods. In this study, approximately 200 genes were targeted. We generated results that were comparable to that of cryosections, and also showed that the workflow could be combined with immunofluorescence staining, highlighting the multi-omics capability of the platform. Scn4b was found to be expressed in the lower cortex and striatum in our mice datasets, which corresponds to published in-situ hybridization (ISH) data as well. We anticipate that the Veranome's reliable, easy-to-use, end-to-end multi-omics solution will facilitate in-situ single-cell transcriptome profiling on many other FFPE sample types and applications.

Extracting CpG methylation from PacBio HiFi whole genome sequencing

Technology Development

POSTER **599**

Richard Hall, **Daniel Portik**, **Kristofor Nyquist**, **Aaron Wenger**

Pacific Biosciences (PacBio), Menlo Park, CA

PacBio HiFi sequencing has been demonstrated to provide the most accurate and complete characterization of human genomes, detecting single-nucleotide variants, indels, and structural variants with high precision and recall using a single sequencing experiment. Here we show that the same experiment also characterizes the epigenome, providing CpG methylation without additional sample prep or sequencing.

PacBio sequencing observes a polymerase in real time as it incorporates fluorescently labeled nucleotides to synthesize a DNA strand. The labels identify the nucleotide sequence. The kinetic signatures – pulse width (time of incorporation) and inter-pulse duration (time between adjacent incorporations) – correlate with chemical modifications to the canonical DNA bases. Some modifications, like N6 methyladenine, have strong kinetic signatures that allow accurate detection with single observations. Others, like the 5-methylcytosine (5mC) modification present in human genomes, have a lower signal-to-noise ratio, which have necessitated high sequence coverage, specialized sample prep, or averaging over genomic regions to achieve high accuracy. PacBio HiFi sequencing, which provides multiple serial observations of the same molecule, opens new methodological approaches. We implemented a multilayer convolution neural network that combines kinetic measurements from the multiple passes and assumes symmetric methylation of CpG sites, which is typical in mammals. We trained the model on fully unmethylated (whole-genome amplification) and fully methylated (M.SssI-treated) HiFi reads. In that context, the network distinguishes methylated from unmethylated CpG in single molecules with an accuracy of around 80% at typical HiFi reads lengths (15 kb).

We further applied the approach on 3 samples from the Genome in a Bottle (GIAB) reference materials, for which CpG methylation is available from other approaches, including bisulfite sequencing, nanopore sequencing, and methylation microarrays. The PacBio HiFi CpG methylation calls have high correlation with the orthogonal approaches. By simultaneously providing accurate SNV calling and phasing, the HiFi reads reveal haplotype-specific methylation including parental imprinting. This complete view of the genome and epigenome in a single experiment provides the opportunity to understand new aspects of evolution, disease, and diversity

Combining highly multiplexed protein and transcriptome-wide RNA expression to achieve spatially resolved multiomic characterization of the tumor microenvironment in FFPE tissues

Technology Development

POSTER **600**

Marlon Stoeckius, Jennifer Chew, Naishitha Anaparthy, Jun D. Chiang, Christina Galonska, Karthik Ganapathy, Ryo Hatori, Alexander Hermes, Layla Katiraee, Anna-Maria Katsori, William Nitsch, Patrick Roelli, Joe Shuga, Rapolas Spalinskas, Mesruh Turkekul, Cedric R. Uytingco, Benton Veire, Dan Walker, Neil I. Weisenfeld, Stephen Williams, Zachary Bent

10x Genomics, USA, Pleasanton, CA

The relationship between cells and surrounding tissue is critical to our understanding of developmental biology and disease progression. Traditionally, Hematoxylin and Eosin (H&E) staining has been used to annotate and characterize tissues and associated pathologies. Recent single analyte approaches spatially interrogate targeted or transcriptome-wide expression of RNA in tissue sections, while others capture phenotypes using a limited number of protein markers. However, for a more comprehensive understanding of the unique characteristics of cell types, cell states, and cell-cell interactions in tissues, multiple layers of information are needed and must be studied together.

Here we demonstrate a novel, streamlined multiomic spatial assay that integrates histological staining and imaging with simultaneous transcriptome-wide gene expression and highly multiplexed protein expression profiling from the same formalin-fixed paraffin embedded (FFPE) tissue section at high resolution. In short, tissue sections from archived FFPE samples are placed on slides with arrayed capture oligos with unique positional barcodes. The H&E or immunofluorescence stained tissue is then imaged, followed by incubation with transcriptome-wide probes and a high-plex DNA-barcoded antibody panel containing intra- and extracellular markers. Transcriptome probes and antibody-barcodes are then spatially captured on the slide and converted into sequencing-ready libraries. Our data analysis and interactive visualization software enable interrogation of all data layers (H&E/immunofluorescence, RNA, protein) from the same tissue section.

We apply this method to simultaneously measure gene and protein expression within the tumor microenvironment of archived human breast cancer and melanoma FFPE samples using whole transcriptome probes and an immune-oncology antibody panel. Taken together, our data demonstrates that a spatially resolved, multiomic approach provides a more comprehensive understanding of cellular behavior in and around tumors, providing new insights into disease progression, predictive biomarkers, drug response and resistance, and development of therapeutic targets.

Novel platform for the simultaneous phenotyping and genotyping of hundreds of thousands of individual cells and cell-to-cell interactions

Technology Development

POSTER **601**

Magali Soumillon[1], Sahbra Eldosougi[1], Nicolas Chevrier[1,2]

1 Flexomics LLC, Boston, MA
2 Pritzker School of Molecular Engineering, the University of Chicago, Chicago, IL

Single cell genomics, by enabling cell-by-cell molecular characterization, has become an essential tool for the analysis of complex cell populations. However, complementing single cell genomics with live cell assays, a critical component to ensure functional relevance, remains a challenge.

Here we present a novel high-throughput approach bridging genomics and functional analysis by enabling large-scale simultaneous phenotyping and genotyping of individual cells and cell-to-cell interactions. Hundreds of thousands of cells can be isolated, cultured, assayed, imaged, and annotated on our proprietary picowell array. Once the phenotypic data has been collected, genetic material is captured on our pre-loaded barcoded beads for downstream single cell genomic analysis. The key feature of our platform resides in our ability to directly link the resulting single cell genomic data to the functional assay performed on individual cells thanks to our 10 million optically readable barcodes that are precisely mapped to each individual cell or pair of cells onto our array. Moreover, our consumable has been specifically designed to allow the co-capture of pairs of cells for the analysis of cell-to-cell interactions, a key requirement for functional analysis such as immune cell response to a specific antigen.

To demonstrate the capabilities of our platform, we focused on the identification and characterization of neoantigen-specific T-Cell Receptors (TCRs). We co-captured autologous pairs of Antigen Presenting Cells (APCs) and CD8+ primary T-cells in our picowell array and identified T-cells secreting cytokines in response to the antigen presentation via miniaturized gold-standard fluorescent immunoassays. RNA from the captured cells was then barcoded in each individual picowell and our targeted single-cell RNA-seq workflow allowed us to collect expression data for a subset of 20 T-cell marker genes as well as retrieve full-length TCR sequences for further clonotype analysis.

With over one million of picowells on a microscope slide format, our platform has the capability to screen hundreds of thousands of single cells or cell-to-cell interactions while seamlessly integrating with existing imaging systems and incorporating gold-standard cell-based fluorescent assays. By coupling live cell analysis and single cell genomics in a high-throughput manner, our platform opens the door to the next generation of single cell analysis to empower a new era of research discovery.

Combined ResolveDNATM whole-genome amplification with Twist Human Core Exome panel to generate high-quality human whole-exome data from single cells

Technology Development

POSTER 602

Swetha D. Velivela[1], **Jon S. Zawistowski**[1], **Isai Salas-González**[1], **Viren Amin**[1], **Victor J. Weigman**[1], **Siyuan Chen**[2], **Gary L. Harton**[1], **Jason A.A. West**[1]

1 BioSkryb Genomics, Inc., Durham, NC
2 Twist Bioscience, South San Francisco, CA

Primary template-directed amplification (PTA) provides unprecedented genomic coverage and uniformity from single cells or low-input DNA, with high SNV calling sensitivity and minimal allelic dropout relative to existing methods. While assessment of the whole genome may be critical for some studies of single cells, other studies may aim to look exclusively at coding regions using whole-exome capture panels. This allows for a more cost-effective initial genomic survey of single cells. Here we report our success using the ResolveDNA™ WGA kit, based on PTA technology, combined with the Twist Human Core Exome hybrid capture panel (Twist Bioscience:PN:102025). ResolveDNA™ WGA combined with whole-exome sequencing of single cells will be a powerful tool to obtain high-quality exome data important for drug development & disease screening.

The Twist target enrichment panels yield highly uniform coverage due to the double-stranded nature of their probe set. We employed the Twist Human Core Exome panel (33Mb) to enrich Twist libraries derived from WGA products of GM12878 single cells using the ResolveDNA™ WGA kit. Following WGA and exome capture according to the manufacturers' instructions, we sequenced (2) 8-plex pools on a NovaSeq 6000. The combined single-cell PTA protocol with the TWIST Human Core Exome panel achieved an average of 98% of the bases covered at 1x and 92% of the bases covered at 10X or higher, when downsampled to 40 million paired-end reads per library. We obtained robust coverage uniformity demonstrated by an average Fold 80 base penalty of 2.39, with a tight range between 2.29 and 2.76, as well by GC/AT dropout rate. This resulted in SNV calling sensitivity of 90% and 88% precision for these single cells.

Here we have shown the power of combining uniform single-cell whole genome amplification with whole-exome capture panels to allow reliable ascertainment of the exome, even at single-cell resolution. Coupling of these technologies has far-reaching potential in diverse applications including cancer, neurology, and preimplantation genetic testing.

Magelia, a versatile multi-OMICs platform for full automation and miniaturization of complex molecular biology applications

Technology Development

POSTER 603

Sebastian Aguilar Pierlé[1], **Ouerdia Maskri**[1], **Camille Soucies**[1], **David Gosselin**[1], **Amel Bendali**[1]

1 Inorevia, Paris, France

Multi-OMICs applications are revolutionizing research and medicine. However, upstream sample prep applications are lagging due to scarce biological samples and the requirement for highly trained staff. The need to treat low-input samples paired with an innovative automation approach is thus clear. Magelia, a multi-OMICs platform combines patented technologies for miniaturization/precise bead handling for efficient, low-input, sample preparation. The platform enabled automation of two challenging molecular biology protocols while showcasing outstanding performance: PCR-free library preparation for WGS/typing of clinical samples; and RNA enrichment for virome characterization through depletion.

Whole-genome-based typing methods for pathogenic Salmonella enterica are becoming the standard due to improved resolution/reduced cost. PCR-free WGS library prep was performed blindly, manually and in the Magelia, using a commercial kit, on low-input clinical S. enterica samples. Importantly, Magelia treated samples yielded sufficient sequencing material while challenging the kit's lower limit by 50x. Platform treated samples resulted in consistently higher library yield compared to manual preparation. Furthermore, 20x less adapter-dimers were detected in Magelia treated samples, resulting in higher quality scores and superior depth. This allowed for improved assembly and enabled identification of all of the Magelia treated samples' types versus 62% for manual treatment. Suitable coverage allowed for detection of 1,7x more clinically relevant genetic elements in Magelia treated samples. Finally, a SNP that revealed a bovine origin for a particular isolate was identified in a Magelia treated sample and not in its manually treated counterpart.

One of the main challenges for virome studies is enriching small amounts of viral particles from abundant host/bacterial genomes. Bats are widespread mammals whose biological/ecological characteristics favor pathogen persistence and spread, including important zoonotic viruses that cause disease in humans and domestic animals. To gain knowledge on the viral community of a widely distributed bat species, we characterized the guano virome of an endemic Chilean bat: Myotis chiloensis. By depleting total RNA in Magelia, we were able to secure, to our knowledge, the highest proportion of viral reads for a bat guano virome study to date. The resolution achieved here enabled us to identify/assemble a new bat alphacoronavirus closely related to PEDV, an important viral pathogen with high mortality rates in suckling piglets.

PCR-free WGS for low-input Magelia-treated samples enabled characterization of clinical isolates to the single nucleotide resolution. Improved RNA enrichment in the platform unlocked unprecedented resolution/depth for a bat virome study, empowering epidemiological surveillance.

The NHGRI Genome Technology Development Coordinating Center

Technology Development

POSTER 604

Mark D. Adams, **Brent Berwin, Gregg TeHennepe, Brianne Angello, Qihui Zhu**

The Jackson Laboratory, Bar Harbor, ME

The National Human Genome Research Institute (NHGRI) Genome Technology (GT) program supports over 50 research groups in academia and industry devoted to developing new technologies for studying the structure and function of the human genome. In 2021, the NHGRI Genome Technology Development Coordinating Center (TDCC) was launched at The Jackson Laboratory to support these diverse grants and expand their impact. Grantees in the areas of nucleic acid sequencing technology, nucleic acid synthesis technology, and technology development for genomic research and clinical applications are conducting a range of cutting edge research leading to new ways to decode natural DNA and RNA sequences, including base modifications, developing novel chemistry and new devices, and exploring ways to understand the dynamic structure and activity of the genome in individual cells. The goal of the TDCC is to create a working consortium of independently-funded investigators based on their shared interests in advancing the technologies used for characterizing the human genome and driven by a desire to support broad implementation of their approaches by the scientific community. The TDCC has three primary objectives: 1) to encourage and support collaboration among grantees, 2) to promote standards in genomic technologies, and 3) to enhance the impact of the GT program through outreach to the scientific community to support adoption of these powerful new approaches to genome characterization to address human health and disease. Current activities are related to exploring gaps and needs related to quality metrics and benchmarking approaches for selected technologies, materials, and metrics for selected technologies. The TDCC also conducts a range of outreach activities to highlight the work of the GT program and encourages new users of these innovative technologies. These include promoting access to methods, materials, and publications, through the TDCC website (www.genometdcc.org), hosting an annual grantee and open meetings, and organizing a series of technology area forums that focus on particular application areas. In this presentation, we will review the discussions of genomic standards coordinated by the TDCC, share recent advances of the GT program, and seek ideas and input from the AGBT community on priorities for the TDCC's outreach efforts.

Long amplicon phasing by Transposase-Induced Multiplex-Encoded Read sequencing (TIMER-seq)

Technology Development

POSTER 606

Kuo-Chan Hung, **Danny Yun, Jess Smith, Joe Mellor***

seqWell, Inc., Beverly, MA
*Presenting author

The ability to derive information from individual long molecules is a key need for a wide range of applications ranging from genome phasing (haplotyping) and metagenomic deconvolution to mRNA isoform detection, synthetic biology, and beyond. While long-read sequencing methods currently represent gold-standard approaches to many of these applications, long-read platforms often present other shortcomings: in scalability (ease of analyzing large numbers of samples), as well as sensitivity (resolution of very large numbers of individual molecules). Applications that require high sensitivity and scalability therefore often depend on short-read sequencing platforms which provide an ability to sequence orders of magnitude more sample fragments.

Sequencing of nucleic acid fragments in the length range of 1kbp to 20kbp represents a particularly challenging subset of applications; this so-called "valley of death" lies above the length limitation of short read sequencers but below the ideal fragment length range (> 20kbp) of long read sequencing. To address this gap between short and long-read platforms we developed a novel method that permits phasing of long fragments using short reads that also takes advantage of the multiplexing and fragment counting capacity of high-throughput short read Illumina sequencers. This method, called transposase-induced multiplex-encoded read sequencing (TIMER-seq), is designed to address the challenge of phasing multiple samples in a highly multiplexed format that is useful for various applications.

TIMER-seq utilizes a novel transposase-based workflow to first tag individual samples with sample-specific barcodes and then barcode pooled mixtures of these tagged samples with "phase barcodes" that add the same barcode at multiple locations of individual long fragments. After sequencing, fragments carrying the same phase barcode are used to recover the phasing information from individual molecules, with the additional benefit of high levels of multiplexing of from the sample-specific barcodes. The pooled format of the phase barcoding process presents significant workflow advantages for processing large numbers of samples with an efficient and scalable workflow.

In this study, we demonstrate proof-of-concept data for generating phasing information from multiplexed 6kb amplicons derived from human genomic DNA using TIMER-seq. We anticipate the TIMER-seq method will have significant potential to be applied to single cell sequencing, multiplexing screening in genetic diseases and pathogens, bacterial and viral metagenomics, agriculture, and beyond.

Technology Development

POSTER 607

CRISPR-targeted nanopore sequencing generates near Megabase-scale reads for identifying cancer gene fusions

GiWon Shin[1], **Bo Zhou**[2], **Stephanie U. Greer**[1], **Hanlee P. Ji**[1]

1 Division of Oncology, Stanford University School of Medicine, Stanford, CA
2 Department of Psychiatry and Behavioral Sciences, Department of Genetics, Stanford University School of Medicine, Stanford, CA

Gene fusion is one of the major genetic causes of human cancers. Genome sequencing approaches to discover gene fusions have performance limitations because of the variable breakpoints located throughout a large region of the genome. Moreover, these methods, regardless of RNA- and DNA-based, do not detect or fully characterize complex structural rearrangements that include inversions and they frequently lack the sensitivity to identify sub-clonal fusions with low allelic fractions. As a consequence, a little is known for the clinical implication of the patient-specific breakpoint locations or complexity of the fusion events.

Here, we present a novel method that generates sub-Megabase-long sequencing reads that includes cancer gene fusions, by which the exact breakpoint can be determined with a base pair resolution, and other complex events in phase with the fusion can also be determined simultaneously. Our method uses in vitro CRISPR-targeting and electrophoresis-based size selection to enrich high molecular weight fragments with the target events, and sequences the fragments with the Oxford Nanopore sequencing. We developed an optimized guide RNA design process where we assess candidates for their uniqueness among GRCh38 k-mers and their off-targets predicted by considering the number and location of mismatches. Moreover, we have optimized the experimental procedure by testing different electrophoresis conditions, and different library preparation methods.

We are using as a test case MLL (or KMT2A) fusions, which is frequent in pediatric acute lymphoblastic leukemias. Our sequencing approach evaluates six 0.3 Mb genomic segments that include the MLL gene itself as well as five frequent translocation partner genes. Using leukemia cells collected from patient peripheral blood, we extract, enrich and sequence targets. We provide a demonstration of rapidly identifying translocations with nanopore sequencing that match clinical cytogenetic results. Importantly, we show how this method can resolve rearrangements that include a mixture of structural variants such as inversions and other alterations.

Index

Index

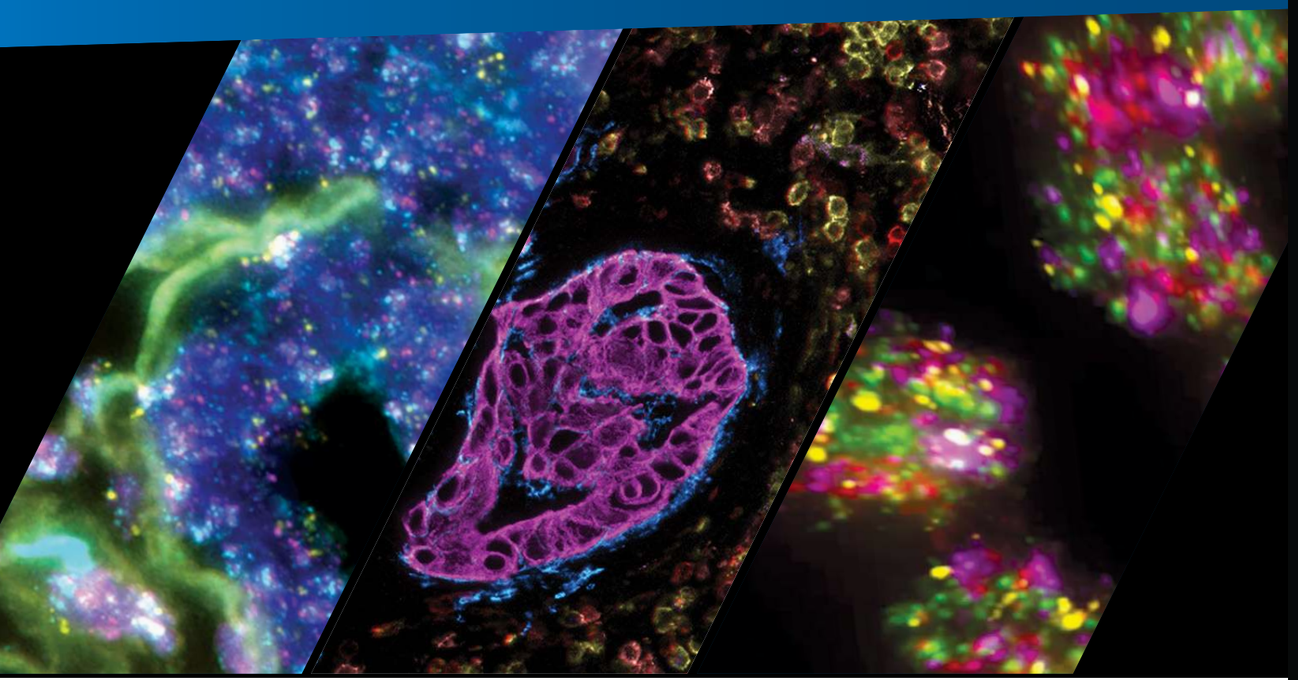
Index

Index

V	Vallebueno-Estrada	Miguel	Gregor Mendel Institute, Austrian Academy of Sciences	239
	VanSchoiack	Alison	NanoString® Technologies	103
	van Velthoven	Cindy	Allen Institute for Brain Science	110
	Velivela	Swetha	BioSkryb Genomics, Inc	237
W	Walter	Dagmar	10x Genomics	166
	Wenger	Aaron	Pacific Biosciences	199
	Williams	Louise	New England Biolabs	41
	Williams	Sion	Sylvester Comprehensive Cancer Center; University of Miami Miller School of Medicine	42
	Wu	Lidan	NanoString Technologies	219
X	Xiao	Chunlin	National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health	77
	Xu	Yi	Zymo Research Corporation	207
Y	Yang	Rendong	The Hormel Institute, University of Minnesota Masonic Cancer Center, University of Minnesota	39
	Yayon	Nadav	Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton	215
Z	Zabrecky	Alita	Dolomite Bio	167
	Zhang	Peng	Center for Inherited Disease Research (CIDR), Department of Genetic Medicine, Johns Hopkins University School of Medicine	63
	Zhang	Yanping	Interdisciplinary Center for Biotechnology Research, University of Florida	174
	Zhang	Yuji	Department of Epidemiology and Public Health, University of Maryland School of Medicine	123
	Zheng	Pengyu	Department of Physics, Northeastern University Department of Biological Regulation, Weizmann Institute of Science	176
	Zhou	Jing	Singleron Biotechnologies, Inc.	169
	Zhu	Wenqi	Singleron Biotechnologies	197
	Zimmerman	Stephanie	NanoString® Technologies, Inc	122

Performance-Leading Innovation in Spatial and Single-Cell Omics

BRUKER



Driving Single-Cell and Subcellar Discovery Across the Omics Spectrum

Rapid technological advances have enabled progressively sophisticated and robust genomic, transcriptomic, and proteomic analyses. Through a range of innovative technologies, Bruker is opening new avenues of inquiry to help researchers gain greater multi-omic understanding across spatial biology applications. The **Acuity Spatial Genomics** product suite offers genome-wide *in situ* visualization of spatially resolved 3D chromatin architecture in individual cells and cell populations. **Canopy Biosciences' ChipCytometry** provides the ability to perform highly multiplexed targeted spatial and single-cell proteomics. Bruker's **Vutara VXL Single-Molecule Localization** bioimaging workstation supports flexible, multiplexed omics research beyond the diffraction limit, and the latest Bruker **timsTOF SCP Mass Spectrometry** instruments now enable routine *ex situ* unbiased proteomics of single cells.

Acuity SPATIAL GENOMICS

A BRUKER COMPANY CANOPY BIOSCIENCES

Vutara VXL

timsTOF SCP

Visit www.bruker.com/**AGBT**

WE HAVE A MESSAGE FROM YOUR FUTURE SELF

IT'S WAITING AT THE COLLIER SUITE.

DON'T MISS OUR
AGBT SPONSOR WORKSHOP
WEDNESDAY, JUNE 8
AT EXACTLY 4:27 PM

bionano GENOMICS