AGBT2024™

February 5-8 • Orlando, FL

ULTIMA GENOMICS

Tuesday, February 6 / 12 PM / Grand SIerra D & E

Genomics Unleashed: The $100 Genome and Beyond

Workshop Summary

The UG 100™ is here to unleash genomics at scale. Learn how our collaborators are shifting the paradigm; from panels to genomes, from late-stage to early detection, and from narrow to large-scale exploration.

Stacey Gabriel, Ph.D.
Broad Institute of MIT and Harvard

Dan Landau, M.D., Ph.D.
Weill Cornell Medical College
New York Genome Center
New York University

Aziz Al'Khafaji, Ph.D.
Broad Institute of MIT and Harvard

Rahul Satija, D.Phil
New York Genome Center
New York University

Will Salerno, Ph.D.
Regeneron Pharmaceuticals

Technical Breakfast Workshops

Wednesday, February 7 / 8 AM - 9 AM
Ultima Genomics Suite, Grand Sierra C

Introducing a New Era: Part Per Million Detection Accuracy with ppmSeq™ Technology

Discover how much more you can find with one-in-a-million accuracy using Ultima's ppmSeq™ technology. Combine extreme, industry leading SNV accuracy with sequencing at scale to reach unprecedented resolution.

Thursday, February 8 / 8 AM - 9 AM
Ultima Genomics Suite, Grand Sierra C

Introducing the Next Wave of Single Cell and Omics at Scale

Access economies of scale with the UG 100™ to transform the breadth and depth of your exploration. Join us for a deep dive on how to scale your experiments and amplify your discovery.

AGBT2024™

GENERAL MEETING

Caribe Royale Orlando • February 5-8, 2024

It is our honor to welcome you to the **Advances in Genome Biology and Technology (AGBT) General Meeting 2024.**

Now in its 24th year, AGBT is recognized as a premier gathering for the genomic research community.

Over the next few days, we will focus on collaboration, dialogue, and innovation 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 renowned conferences and networking events.

Upcoming meetings include:
AGBT Ag (April 15-17, Phoenix, Arizona) and AGBT Precision Health (September 4-6, Denver, Colorado)

AGBT2024™
AGRICULTURE

REGISTER NOW

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

Organizing Committee

We are incredibly grateful to our meeting Co-Chairs Eric Green, Len Pennacchio, Beth Shapiro, and the entire Scientific Organizing Committee for the countless hours they invested in making AGBT24 a content-rich experience.

Penelope Bonnen
Baylor College of Medicine

Federica Di-Palma
Genome British Columbia

Kim Doheny
Johns Hopkins University

Stacey Gabriel
Broad Institute of MIT and Harvard

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

Martin Hirst
University of British Columbia

Elinor Karlsson
Broad Institute, MIT and Harvard

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

AGBT2024™

We gratefully acknowledge the generous support provided by the following companies

GOLD

ULTIMA GENOMICS

SILVER

Element Biosciences

10x GENOMICS

Complete GENOMICS™

BRONZE

SCALE biosciences

BioSkryb GENOMICS

TWIST BIOSCIENCE

NEW ENGLAND Biolabs

nanoString

SINGULAR GENOMICS

WATCHMAKER GENOMICS

PacBio

CONTRIBUTING SPONSORS

illumina

IDT INTEGRATED DNA TECHNOLOGIES

Roche

BRUKER

Volta Labs

TECAN

Lunaphore a biotechne brand

Oxford NANOPORE Technologies

CHRYSALIS BIOMEDICAL ADVISORS

Agilent

Nabsys

Deepcell

ACD a biotechne brand

PIXELGEN TECHNOLOGIES

Moleculent

ROSALIND DATA MADE HUMAN

vizgen

revvity

Dxome

General Information

Registration / Hospitality Desk

Sunday, February 4	12:00 p.m. – 8:00 p.m.
Monday, February 5	8:00 a.m. – 8:00 p.m.
Tuesday, February 6	7:30 a.m. – 9:30 p.m.
Wednesday, February 7	7:30 a.m. – 8:00 p.m.
Thursday, February 8	7:30 a.m. – 6:00 p.m.

Name Badges

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

Transportation

Orlando International Airport (MCO) is 16 miles from Caribe Royale. The transfer time to the resort is approximately 20 minutes, depending on traffic. Shuttles will depart every 30 minutes at the top of the hour and half-past the hour during the following times:

Arrival Shuttles:

Sunday, February 4	1:00 p.m. – 8:00 p.m.
Monday, February 5	7:00 a.m. – 4:00 p.m.

Departure Shuttles:

Friday, February 9	4:00 a.m. – 2:00 p.m.

Social Events

Welcome Dinner – **Monday, February 5 - 7:15 p.m. – 10:00 p.m.**, Poolside
Women's Networking Event – **Tuesday, February 6 – 5:15 p.m. – 7:15 p.m.**, Sun Deck
Sponsor's Social (Connect, Collaborate, and Chill) – **Tuesday, February 6, beginning at 9:30 p.m.** at Sponsor Promenade
AGBT Dinner – **Wednesday, February 7 – 6:10 p.m. – 7:25 p.m.**, The Grove
Farewell Dinner (Mardi Gras) – **Thursday, February 8 - 7:00 p.m. – Midnight**, The Grove

Conference Technology

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

Start Posting!
#AGBT24

Future Unveiled

Element Biosciences invites you to our workshop and suite that promises to redefine the limits of what's possible in sequencing today, and in the future. Brace yourself for an experience where cutting-edge technology meets scientific brilliance.

ABC and AVITI™—Opening a gateway to the next evolution of sequencing

Grand Sierra Ballroom D & E
Wednesday, February 7
12:05 – 1:05pm

Element Suite:
Grand Sierra Ballroom F

Element Biosciences

Agenda

Sunday, February 4

1:00 p.m. – 8:00 p.m.	**Transfers from Orlando International Airport (MCO)**
12:00 p.m. – 8:00 p.m.	**Meeting Registration / Hospitality Desk** Palms Registration

Monday, February 5

7:00 a.m. – 4:00 p.m.	**Transfers from Orlando International Airport (MCO)**
8:00 a.m. – 8:00 p.m.	**Meeting Registration / Hospitality Desk** Palms Registration
	Scientific Program
Plenary Session:	**Genomics I (Eric Green, National Human Genome Research Institute, Chair)** Palms Ballroom II & III
5:00 p.m. – 5:10 p.m.	**Opening Remarks**
5:10 p.m. – 5:40 p.m.	**Stephanie Hicks, Johns Hopkins** "Scalable computational methods and software for single-cell and spatial data science"
5:40 p.m. – 6:10 p.m.	**Anshul Kundaje, Stanford University** “Debugging, denoising, debiasing and interpreting the causal sequence syntax of regulatory profiling experiments with efficient deep learning models”
6:10 p.m. – 6:40 p.m.	**Angela Brooks, University of California, Santa Cruz** "Systematic assessment of long-read RNA-seq methods for transcript identification"
6:40 p.m. – 7:00 p.m.	**Sadie VanHorn, (Abstract Selected) Washington University, St. Louis** “High-resolution single-cell lineage tracing at scale reveals reprogramming dynamics”
7:00 p.m. – 10:00 p.m.	**Welcome Reception** Poolside

Tuesday, February 6

	(Morning)
7:30 a.m. – 9:00 a.m.	**Breakfast** The Grove

7:30 a.m. – 9:30 p.m.	**Meeting Registration / Hospitality Desk** Palms Registration
Plenary Session:	**Evolutionary Genomics (Beth Shapiro, University of California, Santa Cruz, Chair)** Palms Ballroom II & III
9:00 a.m. – 9:30 a.m.	**Kirsten Bos, Max Planck Institute** "Genomic reconstructions of ancient pathogens"
9:30 a.m. – 10:00 a.m.	**Bridgett vonHoldt, Princeton University** “Ghost wolves: Reviving an endangered canid species”
10:00 a.m. – 10:30 a.m.	**Sean Eddy, Harvard University** "Computational analysis of RNA structure and function"
10:30 a.m. – 11:00 a.m.	**Coffee Break** Sponsor Promenade
11:05 a.m. – 11:25 a.m.	**Julian Hess, (Abstract Selected) Broad Institute of MIT and Harvard** “High accuracy somatic variant calling on Ultima Genomics facilitates detailed phylogenetic reconstruction across 303 metastatic tumor samples”
11:25 a.m. – 11:45 a.m.	**Poster Flash Talks**
Tuesday, February 6	**(Afternoon)**
12:00 p.m. – 1:30 p.m.	**Gold Sponsor Workshop & Lunch - Ultima Genomics** Grand Sierra Ballroom D&E **Gilad Almogy, Ph.D., Chief Executive Officer, Ultima Genomics** **Mirna Jarosz, Ph.D., Head of Product, Ultima Genomics,** **Stacey Gabriel, Ph.D., Executive Vice President of Platforms and Scientific Execution, Institute Scientist, Broad Institute of MIT and Harvard** **Aziz Al'Khafaji, Ph.D., Associate Director, Technology Development - Genomics Platform, Broad Institute of MIT and Harvard** **Dan Landau, M.D., Ph.D., Professor of Physiology and Biophysics, Weill Cornell Medical College** **Rahul Satija, D.Phil., Core Faculty Member, New York Genome Center** **Will Salerno, Ph.D., Executive Director of Genome Informatics & Data Engineering, Regeneron Pharmaceuticals** "Genomics unleashed: the $100 genome and beyond"

12:00 p.m. – 1:30 p.m.	**AGBT Lunch** The Grove
1:30 p.m. – 3:30 p.m.	**Poster Session with Coffee & Dessert** Palms Ballroom I
Plenary Session:	**Biology (Penelope Bonnen, Baylor College of Medicine, Chair)** Palms Ballroom II & III
3:30 p.m. – 4:00 p.m.	**Mark Blaxter, Wellcome Sanger Institute** "Sequence locally, think globally: biodiversity genomics across the tree of life"
4:00 p.m. – 4:20 p.m.	**Sarah Bates, (Abstract Selected) National Human Genome Research Institute/NIH** "The power and importance of inclusive, fact-driven communications in human genomics"
4:20 p.m. – 4:40 p.m.	**Jiwoon Park, (Abstract Selected) Weill Cornell Medicine** "The Spatial Atlas of Human Anatomy (SAHA) project: unveiling cellular landscapes and orchestrating a new paradigm in precision medicine"
4:40 p.m. – 5:00 p.m.	**Meredith Wright, (Abstract Selected) Rady Children's Institute for Genomic Medicine** "Leveraging population databases and machine learning models to scale whole genome newborn sequencing"
Tuesday, February 6	**(Evening)**
5:15 p.m. – 7:15 p.m.	**Dinner On Your Own**
5:15 p.m. – 7:15 p.m.	**Women's Networking Event-Featuring Astronaut Kathleen "Kate" Rubins** Sun Deck
Concurrent Session:	
7:30 p.m. – 9:30 p.m.	**Cancer (John McPherson, University of California, Davis, Chair)** Grand Sierra Ballroom D&E

7:30 p.m. – 7:50 p.m.	**Ruli Gao, Northwestern University** "Studying isoforms and adaptive evolution of brain metastasis with long read single cell RNA sequencing"
7:50 p.m. – 8:10 p.m.	**Cynthia Lawley, Olink** "Next gen proteomics combined with genomics identifies potential therapeutic targets for breast cancer"
8:10 p.m. – 8:30 p.m.	**Jessica Nordlund, SciLifeLab, Uppsala University** "Mapping ex vivo drug responses in single cells"
8:30 p.m. – 8:50 p.m.	**Brian Haas, Broad Institute, Methods Development Laboratory** "High Accuracy Fusion Transcript Detection from Long-Read Isoform Sequencing using CTAT-LR-Fusion"
8:50 p.m. – 9:10 p.m.	**Holger Heyn, Centro Nacional de Análisis Genómico (CNAG)** "Spatio-temporal dissection of colorectal cancer initiation using whole transcriptome imaging"
9:10 p.m. – 9:30 p.m.	**Hanghui Ye, The University of Texas MD Anderson Cancer Center** "A pan-cancer single-cell analysis of intratumoral copy number diversity and evolution"
Concurrent Session:	
7:30 p.m. – 9:30 p.m.	**Computational Biology (Mike Zody, New York Genome Center, Chair)** Palms Ballroom II & III
7:30 p.m. – 7:50 p.m.	**Ana Conesa, Spanish National Research Council** "Quantifying isoform expression and transcriptome complexity by high-throughput single-molecule sequencing"
7:50 p.m. – 8:10 p.m.	**Javan Okendo, National Institute of Health** "Telomere-to-telomere genome assemblies for commonly used zebrafish lab strains"
8:10 p.m. – 8:30 p.m	**Maria Sierra, Weill Cornell Medicine** "Detection of human herpesviruses in Alzheimer's brains and implications for neurogenerative diseases"

8:30 p.m. – 8:50 p.m	**Yueyao Gao, Broad Institute of MIT and Harvard** “T2T-ACE: A novel approach for accurate CNV evaluation using Telomere-to-Telomere assemblies”
8:50 p.m. – 9:10 p.m.	**Alexander Hoischen, Radboud University Medical Center** “Assessing HiFi genome sequencing as first-tier test in rare disease genetics”
9:10 p.m. – 9:30 p.m.	**Koen van Gassen, University Medical Center, Utrecht** “Replacing karyotyping and FISH with Hi-C chromosome conformation capture”
Starting at 9:30 p.m.	**Sponsor’s Social (Connect, Collaborate, and Chill)** Sponsor Promenade
Wednesday, February 7	**(Morning)**
7:30 a.m. – 9:00 a.m.	**Breakfast** The Grove
7:30 a.m. – 8:00 p.m.	**Hospitality Desk** Palms Registration
Plenary Session:	**Genetics (Stephen Montgomery, Stanford University School of Medicine, Chair)** Palms Ballroom II & III
9:00 a.m. – 9:30 a.m.	**Melissa Boneta Davis, Weill Cornell Medicine** "Ancestry in genomics of cancer disparities"
9:30 a.m. – 10:00 a.m.	**Melina Claussnitzer, Broad Institute, Massachusetts General Hospital, Harvard Medical School** "From metabolic disease maps to mechanism to medicines"
10:00 a.m. – 10:30 a.m.	**Federica Di Palma, Genome British Columbia** "Harnessing responsible genomics for biodiversity: Protecting Colombia's natural wealth"
10:30 a.m. – 11:00 a.m.	**Coffee Break** Sponsor Promenade
11:00 a.m. – 11:20 a.m.	**Konrad Karczewski, (Abstract Selected) Massachusetts General Hospital** “All by All of Us: common and rare variant association testing in 250,000 whole genomes across diverse ancestry groups”

11:20 a.m. – 11:40 a.m.	**Next Gen Leadership Awards (Eric Green, NHGRI, Elinor Karlsson, Broad Institute, MIT and Harvard, Beth Shapiro, University California, Santa Cruz, Co-Chairs)**
Wednesday, February 7	**(Afternoon)**
12:00 p.m. – 2:15 p.m.	**Silver Sponsor Workshops and Lunch** Grand Sierra Ballroom D&E
12:05 p.m. – 1:05 p.m.	**Silver 1 Sponsor Workshop and Lunch – Element Biosciences** **Molly He, Element Biosciences** "ABC and AVITI: Opening a gateway to the next evolution of sequencing"
1:10 p.m. – 2:10 p.m.	**Silver 2 Sponsor Workshop and Lunch – 10X Genomics** **Sarah Taylor, Ben Hindson, Mike Schnall-Levin and Patrick Marks** "Mind blowing science enabled by industry-leading single cell & spatial technologies"
12:00 p.m. – 2:10 p.m.	**AGBT Lunch** The Grove
2:15 p.m. – 4:44 p.m.	**Bronze Sponsor Workshops** **(Kim Doheny, Johns Hopkins University, Chair)** Palms Ballroom II & III
2:20 p.m. – 2:40 p.m.	**Bronze 1 Sponsor - Scale BioSciences** **Giovanna Prout, CEO** "The single cell era, scaled"
2:40 p.m. – 3:00 p.m.	**Bronze 2 Sponsor - BioSkryb Genomics** **Jay West** "The next generation of single-cell technology – resolve more with BioSkryb"
3:00 p.m. – 3:20 p.m.	**Bronze 3 Sponsor - Twist Bioscience** **Emily Leproust, Ph.D. CEO & co-founder of Twist Bioscience** "Unveiling next-gen solutions for precision diagnostics and beyond"

3:20 p.m. – 3:35 p.m.	**Bronze 4 Sponsor - New England Biolabs** **Keerthana Krishnan, Ph.D., Development Group Leader, Next Generation Sequencing** "NEBNext product development: streamlining workflows and addressing new applications"
3:35 p.m. – 3:50 p.m.	**Bronze 5 Sponsor - NanoString** **Joseph Beechem, Ph.D, Chief Scientific Officer and Senior Vice President of Research and Development, NanoString** “NanoString spatial biology roadmap: the holy grail of spatial is here”
3:50 p.m. – 4:05 p.m.	**Bronze 6 Sponsor - Singular Genomics** **Eli Glezer PhD, CSO and Co-Founder** "In-situ sequencing: enabling spatial biology at scale"
4:05 p.m. – 4:20 p.m.	**Bronze 7 Sponsor - Watchmaker Genomics** **Travis Sanders** “Flipping the script on DNA methylation sequencing: A positive readout to enable multimodal analyses”
4:20 p.m. – 4:32 p.m.	**Bronze 8 Sponsor – PacBio** **Isabelle Thiffault, MSc, PhD, FACMGG, FABMGG. Director of Translational Genetics - Clinical Lab. Genomic Medicine Center, Children's Mercy Research Institute, Children’s Mercy Hospitals and Clinics** "Empowering precision medicine research: the potential paradigm shift of HiFi genome sequencing in first-line clinical tests"
4:32 p.m. – 4:44 p.m.	**Bronze 9 Sponsor - n6** **Dr. Pranav Patel, PhD** “iconPCR: A universal method for up to a 99% improvement in NGS library prep workflows and sequencing quality”
4:45 p.m. – 6:10 p.m.	**Poster Session and Wine & Cheese Reception** Palms I Ballroom

Wednesday, February 7 **(Evening)**

6:10 p.m. – 7:25 p.m.
AGBT Dinner
The Grove

Concurrent Session:
7:30 p.m. – 9:30 p.m.
Technology (Elinor Karlsson, Broad Institute of MIT and Harvard, Chair)
Palms II & III

7:30 p.m. – 7:50 p.m.
Dawn Chen, Broad Institute of MIT and Harvard
"Long-range continuous mutagenesis of endogenous genomes"

7:50 p.m. – 8:10 p.m.
Andrew Adey, Oregon Health & Science University
"Atlas-scale single-cell DNA methylation"

8:10 p.m. – 8:30 p.m.
Elizabeth Neuman, University of California, Davis
"Using MS based ISH Approaches to understand gene expression and its relation to metabolomics"

8:30 p.m. – 8:50 p.m.
Hadas Keren Shaul, Weizmann Institute of Science
"Genomics sandbox: a playground for genomics research and science innovation"

8:50 p.m. – 9:10 p.m.
Joshua Levin, Broad Institute of MIT and Harvard
"Development of experimental and computational methods for allelic imbalance analysis from single-nucleus RNA-seq data"

9:10 p.m. – 9:30 p.m.
Huan Wang, Broad Institute of MIT and Harvard
"Systematic benchmarking of imaging spatial transcriptomics platforms in Formalin-Fixed Paraffin-Embedded samples across healthy and cancerous tissues"

Concurrent Session:
7:30 p.m. – 9:30 p.m.
Biology (Federica Di Palma, Genome British Columbia, Chair)
Grand Sierra Ballroom D&E

7:30 p.m. – 7:50 p.m.
Massimo Delledonne, University of Verona
"Adaptive gene loss shapes the evolution of the Phaseolus vulgaris pan-genome during species range expansion and dual domestications"

7:50 p.m. – 8:10 p.m.	**Darrell Dinwiddie, University of New Mexico HSC** “Sequencing of museum biospecimens for the discovery and genomic characterization of orthohantaviruses”
8:10 p.m. – 8:30 p.m.	**Shruti Iyer, Cold Spring Harbor Laboratory** “Using targeted long-read sequencing to identify genetic and epigenetic changes across different stages of colorectal cancer”
8:30 p.m. – 8:50 p.m.	**Brenda Murdoch, University of Idaho** “Centromere content and organization in cattle and sheep Y chromosomes are considerably different than in the human Y chromosome”
8:50 p.m. – 9:10 p.m.	**Miranda Orr, Wake Forest University School of Medicine** “Ultra-high plex single-cell spatial multi-omic imaging of tau neuropathology in human brain sections”
9:10 p.m. – 9:30 p.m.	**Alexander Urban, Stanford University** “Machine learning analysis of deep whole-genome sequencing data in a large cohort of human brains reveals common presence of somatic, mosaic LINE-1 insertions”
Thursday, February 8	**(Morning)**
7:30 a.m. – 9:00 a.m.	**Breakfast** The Grove
7:30 a.m. – 6:00 p.m.	**Hospitality Desk** Palms Registration
Plenary Session:	**Genomics ll (Stacey Gabriel, Broad Institute of MIT and Harvard, Chair)** Palms Ballroom II & III
9:00 a.m. – 9:30 a.m.	**Kristin Laidre, University of Washington** "A new population of the world's largest land carnivore and why it matters"

9:30 a.m. – 10:00 a.m.	**Michael Schatz, Johns Hopkins University** “BioDIGS: bioDiversity and informatics for genomics scholars”
10:00 a.m. – 10:20 a.m.	**Yichen Si, (Abstract Selected) University of Michigan** “FICTURE: Segmentation-free factor analysis for sub-micron resolution spatial transcriptomics”
10:20 a.m. – 11:00 a.m.	**Coffee Break** Sponsor Promenade
11:05 a.m. – 11:25 a.m.	**Thomas Gingeras, (Abstract Selected) Cold Spring Harbor Laboratory** “The EN-TEx resource of multi-tissue personal epigenomes & variant-impact models”
11:25 a.m. – 11:45 a.m.	**Sek Won Kong, (Abstract Selected) Boston Children’s Hospital** “Discordance between a deep learning model and clinical-grade variant pathogenicity classification in a rare disease cohort”
12:00 p.m. – 1:00 p.m.	**Silver 3 Sponsor Workshop & Lunch – Complete Genomics** Grand Sierra Ballroom D&E **Rade Drmanac, Chief Scientific Officer, Complete Genomics** "Advancing sequencing capacities: exploring DNBSEQ™ technology and complete WGS" **Christopher E. Mason, Ph.D., Professor, Department of Physiology and Biophysics, Weill Cornell Medicine** "Spatial oncology maps and space biology maps" **Ioannis Ragoussis, Professor, Department of Human Genetics - Head of Genome Sciences, McGill Genome Centre** "A cost-effective whole genome sequencing method identifies clinically relevant germline variants in patients recruited in a breast cancer clinic"

12:00 p.m. – 1:30 p.m.	**AGBT Lunch** The Grove
Thursday, February 8	**(Afternoon)**
Plenary Session:	**Technology (Chris Mason, Weill Cornell Medicine, Chair)** Palms Ballroom II & III
1:30 p.m. – 2:00 p.m.	**Kathleen Rubins, NASA**
2:00 p.m. – 2:20 p.m.	**Jagadish Sankaran, (Abstract Selected) Genome Institute of Singapore** "Holistic identification of morphogenomic phenotypes using morphology and expression data optimal clustering (MEDOC)"
2:20 p.m. – 2:40 p.m.	**Xin Jin, (Abstract Selected) Scripps Research** "Massively parallel in vivo Perturb-seq reveals cell type-specific transcriptional networks in cortical development"
2:40 p.m. – 3:00 p.m.	**Billy Lau, (Abstract Selected) Stanford University** "Joint single-molecule fragmentomic and methylation signatures of cell-free DNA for multi-omic cancer detection using nanopore sequencing"
3:00 p.m. – 3:20 p.m.	**Dennis Yuan, (Abstract Selected) New York Genome Center** "High-throughput miniaturized-automated workflow for joint high-depth whole genome and whole transcriptome sequencing of thousands of human primary cells"
3:20 p.m. – 3:40 p.m.	**Daniel Turner, (Abstract Selected) Enhanc3D Genomics** "Interrogating 3-dimensional genomic architecture to reveal novel disease biology"
3:40 p.m. – 4:00 p.m.	**Closing Comments, Meeting Feedback and AGBT'25 Anniversary Announcement**
7:00 p.m. – Midnight	**Farewell Dinner - Mardis Gras** The Grove

YOU  US = MIND BLOWING SCIENCE

Your mission is to make incredible discoveries. Our passion is to make that possible. Realize your single cell ambitions with Chromium's expanding versatility. Achieve your spatial explorations with Visium's HD discovery power. Fuel your spatial single cell pursuits with Xenium's ultra precision.

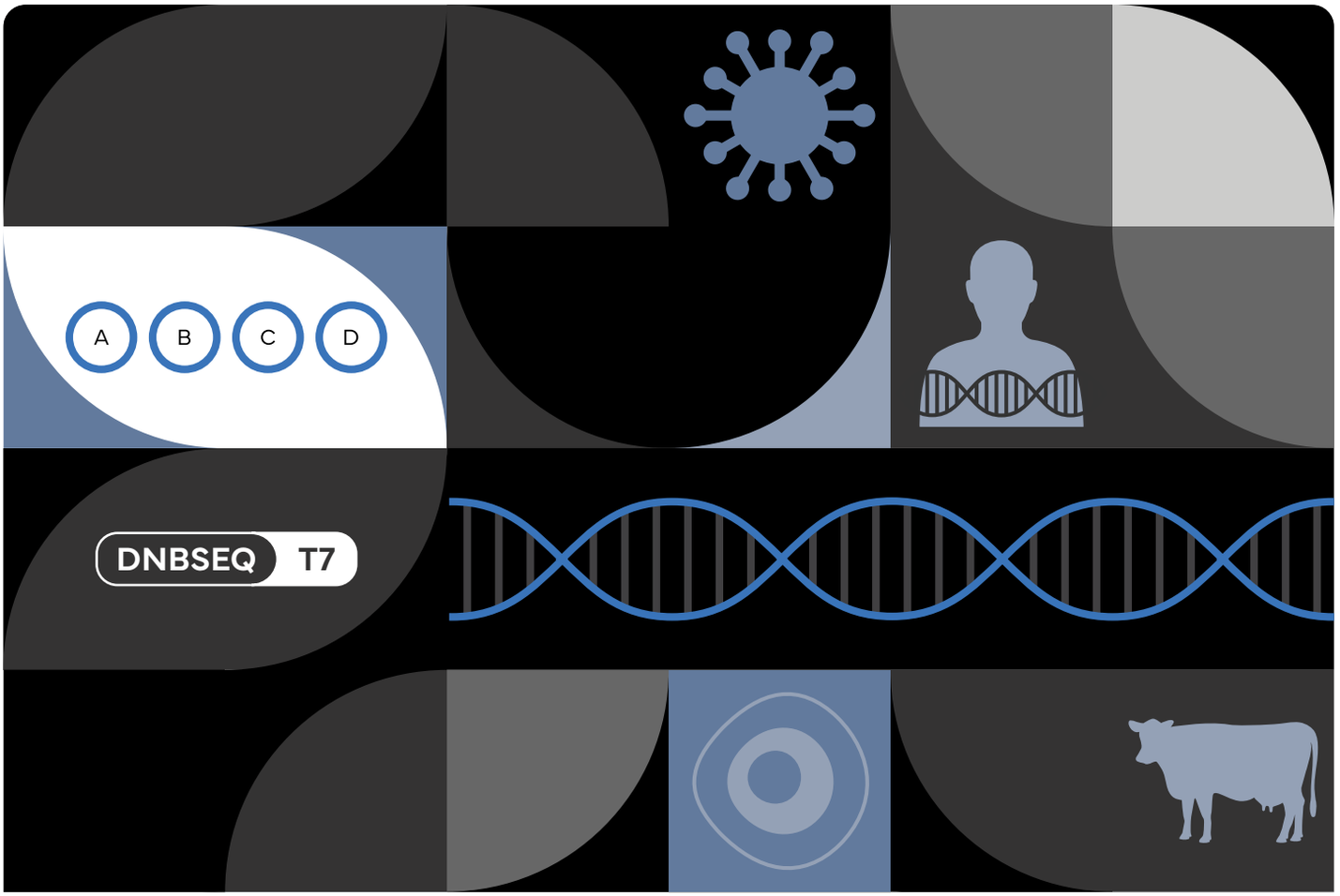
Turn your most visionary research into reality in Bonaire 5

10x GENOMICS

Abstract Table of Contents

page

Complete GENOMICS™



TURN YOUR RESEARCH INTO A MASTERPIECE WITH DNBSEQ-T7

With the flexibility of 4 independent flow cells, the DNBSEQ-T7 Sequencer enables you to create your research masterpiece across a range of applications, from WGS to spatial genomics — all on ONE sequencer, generating up to 20,000 WGS in one year at just $150 per genome.

Get a FREE whole genome test* at our suite: Bonaire 1

Learn more about our Thursday lunch workshop and suite activities
completegenomics.com/agbt24

* This is not associated with or endorsed by AGBT

Cancer Omics

Advancing cancer genomics with cost-effective whole genome sequencing: somatic variant calling using the UG100 platform

Hila Benjamin[1], Ilya Soifer[1], Doron Shem-Tov[1], Maya Levy[1], Islam Oguz Tuncay[2], Baek-Lok Oh[2], Jong-Yeon Shin[2], Young Seok Ju[2], Sangmoon Lee[2], Omer Barad[1], Doron Lipson[1]

1 Ultima Genomics Inc., Newark, CA, USA
2 Genome Insight Inc., San Diego, CA, USA

Somatic variant calling involves the identification of genomic alterations that occur exclusively in somatic cells, such as cancer-associated mutations, requiring deep coverage to enable high sensitivity for low-frequency variants. Characterizing somatic variants across the entire genome therefore significantly benefits from novel cost-efficient sequencing platforms, such as UG100.

DeepVariant is a deep learning-based approach for variant calling which has been successfully applied across many sequencing technologies. We optimized DeepVariant for somatic variant calling using UG100 sequence data from matched tumor-normal sample pairs, improving both variant calling accuracy as well as algorithm running time (up to 10-fold). Deep learning models were trained on mixtures of Genome-In-A-Bottle reference samples with different ratios to simulate somatic variants. Performance on simulated data for VAF>10% showed SNV recall >98% and indel recall >95% with false-positive rate of 0.2/Mb. For VAF range of 5-10%, indel recall was 67% and SNV recall was 86%. To demonstrate the utility of our somatic variant calling, we applied the model to WGS from well characterized cancer cell lines, COLO829 and HCC1395.

Next, we profiled a set of 26 clinical tumor-normal breast cancer sample pairs using WGS (mean tumor coverage: 150x; normal: 70x) and observed high concordance with Illumina sequencing, with >97% of SNVs and >93% of indels detected by Illumina identified also by UG100. UG100 sequencing was deeper and allowed for identification of a larger number of genomic variations, including known driver events in cancer-related genes. The somatic mutational signatures by UG100 were concordant with the Illumina platform, with a cosine similarity of 99.38%.
WGS can provide valuable information on large-scale variation in addition to the small variants. Finally, we demonstrated that optimized open-source and widely utilized pipelines for structural variation and CNV calling (GRIDSS and FREEC) can be used with UG100 data.

In conclusion, our research highlights the utility of UG100 within the field of oncology, demonstrating its capacity for comprehensive and precise somatic variant detection. Unlike targeted panels, which focus on specific genes or regions, WGS provides a much broader view of tumor biology.

Analysis of pancreatic cancer risk variants using long read sequencing

Qiuhui Li[1], Carolina Montano[2], Jessica Hosea[2], Luke Morina[2], Bohan Ni[1], Moira McCormick[3], Justin Paschall[4], Beth Marosy[4], Michelle Kokosinski[4], Jessica Gearhart[4], Brian Craig[4], Alan Scott[4], David Mohr[4], Michelle Mawhinney[4], David McKean[5,6], Nicholas Roberts[5,6], Zhanmo Ni[5,6], Alexis Battle[1,2], Kimberly Doheny[4], Winston Timp[2], Michael Schatz[1], Alison Klein[5,6]

1 Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA
2 Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA
3 Department of Biology, Johns Hopkins University, Baltimore, MD, USA
4 Center for Inherited Disease Research, Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA
5 Department of Oncology, The Sol Goldman Pancreatic Cancer Research Center, The Johns Hopkins Medical Institutions, Baltimore, MD, USA
6 Department of Pathology, The Sol Goldman Pancreatic Cancer Research Center, The Johns Hopkins Medical Institutions, Baltimore, MD, USA.

Pancreatic cancers are the third leading cause of cancer-related deaths in the United States, and the annual number of diagnoses worldwide has more than doubled over the past 25 years. Despite recent improvements in detection and treatment, the 5-year survival rate in the US remains a dismal 12%. Family history is one of the most important known risk factors and increases risk by 2.3 to 32-fold. Association studies leveraging short-read sequencing have identified germline genetic risk factors in BRCA2, PALB2, ATM, and a few other genes. However, only a small fraction (~10%) of familial cancers have recognizable mutations, leaving many patients unaware of their true risk. Comprehensive variation profiling is critical to advancing cancer susceptibility research. However, investigating complex structural variations and certain SNVs - particularly those in highly repetitive regions - remains challenging due to short-read sequencing constraints.

(continue to page 26)

(continued from page 25)

Addressing this need, we are using long-read sequencing to comprehensively assess variants in high-risk familial pancreatic cancer patients. Long-read sequencing produces reads tens of kilobases long that can span difficult-to-assemble regions and improve variation identification. To date, we have performed long-read whole genome sequencing on germline samples from >30 familial cancer patients on the Oxford Nanopore PromethION platform (R10.4 flowcells and LSK114 library prep) to an average mean depth of 58X and n50 of 27.5 kb. We produced highly accurate variant calls using Sniffles2, PAV, and Clair3. Notably, using long reads, we detect SVs and other complex variants in these patient genomes that are not detectable using short reads in genes known to increase risks for hereditary pancreatic cancer and other hereditary cancers. Using SVs cataloged within the families and those identified in other control samples, we are designing an integrative algorithm to detect likely pathogenic germline alterations in patients' genomes. Our ultimate goal is to build a comprehensive variant dataset comprising over 400 familial pancreatic patient samples and unravel the importance of structural variants in cancer. We expect our efforts to substantially improve patient genetic profiling, especially in high-risk families, thereby improving pancreatic cancer prevention and therapy and ultimately saving lives.

Ataxia-telangiectasia mutated (Atm) disruption sensitizes spatially directed H3.3K27M/TP53 diffuse midline gliomas to radiation therapy.

Zachary J. Reitman[1], Avani Mangoli[2], Emily Hocke[3], Vaibhav Jain[3], Michael Aksu[3], Karen Abramson[3], Bronwen E. Foreman[4], Debosir Ghosh[5], María E. Guerra Garcia[1], Harrison Liu[1], Lixia Luo[1], Lore Weidenhammer[1], Nerissa T. Williams[1], Sophie Wu[5], Joshua A. Regal[6], Laura Attardi[7], Oren J. Becher[8], David G. Kirsch[9], David M. Ashley[10], **Simon G. Gregory[11]**

1 Radiation Oncology-Cancer Biology, Duke University;
2 Pediatric Neuro Oncology, Duke University;
3 Duke Molecular Physiology Institute;
4 Duke University School of Medicine;
5 Duke University School of Arts and Sciences;
6 Roswell Park Comprehensive Cancer Center;
7 Stanford University;
8 Pediatric Hematology Oncology, Mount Sinai Health System;
9 Radiation Oncology, University of Toronto;
10 Neuro Oncology, Duke University;
11 Neurosurgery, Duke University

Diffuse midline gliomas (DMGs) are lethal brain tumors characterized by inactivating p53 mutations and oncohistone H3.3K27M mutations that rewire the cellular response to genotoxic stress, presenting therapeutic opportunities. We used RCAS/tv-a retroviruses and Cre recombinase to inactivate p53 and induce K27M in the native H3f3a allele in a lineage- and spatially directed manner, yielding primary mouse DMGs. Disruption of the DNA damage response kinase Ataxia-telangiectasia mutated (Atm) enhanced the efficacy of focal brain irradiation, extending mouse survival. This finding suggests that targeting ATM will enhance the efficacy of radiation therapy for p53-mutant DMG but not p53-wildtype DMG. We used 10xGenomics Xenium spatial in situ sequencing (ISS) platform and an allelic series of primary murine DMG models to identify transactivation-independent p53 activity as the key mediator of such radiosensitivity.

(continue to page 28)

(continued from page 27)

Our ISS data provides the first high-resolution transcriptional analysis at high gene-plexy of a mouse tumor model. High-resolution spatially resolved transcriptomic profiling was especially critical in this model since vasculature and neoplastic compartments within the tumor play distinct roles in therapeutic response. These spatial data showed that irradiation elicited overexpression of Cdkn1a, a key downstream target of p53 that mediates cell senescence and G1-to-S checkpoint arrest, in p53-null tumors suggesting p53-independent mechanisms of Cdkn1a expression. This finding led us to dissect the contribution of p53 transactivation functions that regulate p21 expression to radiosensitization in Atm-null DMGs. These findings implicate transactivation-independent functions of p53 as key determinants of radiosensitivity in Atm-null tumors. Our data provides genetic and mechanistic insight that builds upon studies of pharmacologic ATM inhibition in patient derived xenograft models6. Further studies are needed to determine transactivation-independent mechanisms of p53 and ATM-directed therapies and their impact on overcoming resistance to radiation therapy in patients with H3.3K27M-mutant DMG.

Cellular-resolution multiplexed-CNV FISH captures spatial copy number heterogeneity and clonal architecture of colorectal cancer.

Mei Suen KONG[1], Hui Ting Grace YEO[1], Jiamin TOH[1], Lin Li[2], Jonathan Shaozhong Aow[1], Joanito Irwan IGNASIUS[1], Jen Yi WONG[1], Nirmala ARUL RAYAN[1], Lavanya MUTHUKRISHNAN1, Dominique Camat MACALINAO[3], Wei Qiang LEOW[4], Kiat Hon LIM[4], Bee Huat Iain TAN[3], Kok Hao CHEN[2], Shyam PRABHAKAR[1]

1 Genome Institute of Singapore, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore.
2 Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore.
3 National Cancer Centre, Singapore, Singapore.
4 Department of Anatomical Pathology, Singapore General Hospital, Singapore, Singapore.

In tumors, clonal populations of malignant cells evolve a diverse repertoire of copy number variants (CNVs), often resulting in profound intra-tumor heterogeneity (ITH). ITH is associated with poor prognosis, cancer progression, increased risk of therapy resistance, metastasis and relapse. Traditionally, researchers have studied clonal heterogeneity and tumor evolution using bulk-tissue assays: multi-regional sampling, comparative genomic hybridization microarrays, whole genome sequencing and whole exome sequencing. However, these techniques provide little or no spatial information and have low sensitivity for focal/low copy-number CNVs, particularly when they are subclonal or rare. DNA FISH is the gold standard for detecting CNVs at cellular resolution, but it is limited to one gene per tissue section.

To overcome the above limitations, we developed multiplexed-CNV FISH, a spatial omics assay that detects gain and loss of DNA loci at single-gene (~10kb) and single-cell resolution in human tissues. We validated the accuracy of the assay using cultured cell lines and normal human tissues. We then used CNV-FISH to spatially profile the CN of 35 genes commonly amplified or deleted in the iCMS2 CRC subtype we recently described. In contrast, these loci are mostly diploid in the iCMS3 CRC subtype. We validated CNV-FISH results by inferring CNVs from matched single cell RNA-seq, spatial transcriptomic profiling and WES data, and by examining CNV patterns in data from the TCGA consortium. Remarkably, CNV-FISH analysis of CRC tumor sections identified small and subclonal CNVs not detected by existing methods with resolution of CNV spanning whole chromosome arms.

(continue to page 30)

(continued from page 29)

Our results indicate that CNV-FISH can reveal unexpected and previously undetected CN heterogeneity within tumors, which could potentially underlie treatment resistance and eventual metastasis of cancer cells. The assay could thus serve as a valuable component of cohort-scale cancer omics studies for characterizing frequently mutated loci and driver genes. It will also provide a unique view into malignant cell evolution within a tumor, and potentially even identify rare pre-cancerous CNV clones within nominally healthy tissue. Ultimately, we are optimistic that CNV-FISH and related assays will be adopted in histopathology workflows as a high-resolution tool for cancer diagnosis, patient stratification, and guidance for the selection of therapies.

Characterization of the tumor microenvironment of IDH1 wild type and mutant high grade glioma using targeted protein panels in a spatial context.

Emily Hocke[1], Vaibhav Jain[1], Dmytro Klymyshyn[2], Bassem Ben Cheikh[2], Niyati Jhaveri[2], Michael Prater[3], Nadine Nelson[3], Vinit Pereira[3], Seetha Hariharan[4], Michael Brown[4], Oliver Braubach[2], David Ashley[4], Matthew Waitkus[4], Simon Gregory[1]

1 The Duke Molecular Genomics Core, Duke Molecular Physiology Institute, Durham, United States.
2 Akoya Biosciences, Discovery Applications, MENLO PARK, United States.
3 Abcam, Immuno Oncology, Cambridge, United Kingdom.
4 The Preston Robert Tisch Brain Tumor Center at Duke, Brain Tumor Center, Durham, United States.

Gliomas, a highly destructive cancer, are characterized by unfavorable outcomes even with aggressive treatment such as surgical removal, chemotherapy, and radiation therapy. Mutations in the Isocitrate Dehydrogenase 1 (IDH1) gene are a characteristic feature of Astrocytoma, IDH-mutant tumors, a subtype of glioma that often presents as grade 2/3 tumors and exhibits a propensity to progress to grade 4 tumors at recurrence. IDH1 mutations occur as single amino acid substitutions in the IDH1 enzyme active site, most commonly as IDH1R132H, and leads to the production of D-2-hydroxyglutarate (D-2-HG). D-2-HG is also reported to modulate the tumor microenvironment, but the magnitude of these effects and the mechanisms by which they occur not well understood.

We have used Akoya Biosciences PhenoCode™ Discovery Panels and custom markers to analyze >70 proteins on the PhenoCycler®-Fusion. These proteins allowed us to study immune checkpoints, tissue structure, cellular activation states and tumor landscapes in malignant gliomas with (Astrocytoma, IDH-mutant, n=2) and without (Glioblastoma, IDH-wildtype=2) IDH1 mutations. To achieve this, we used 60 proteins central to the immune system, including immune cell lineages, activation states and checkpoints. We also introduced custom markers to explore metabolism, neuronal elements, and epigenetic factors. This approach leveraged ultra-high-plex spatial phenotyping techniques to identify distinct expression patterns, cell-type profiles, and cell-type spatial distributions in Astrocytoma, IDH-mutant versus IDH-wildtype glioblastoma (GBM) tumors. We identified increased abundance of tumor-infiltrating lymphocytes in IDH-wildtype GBM tumors versus IDH-mutant tumors. These differences were caused in part by a marked increase in T helper cells in IDH-WT GBMs, as well as a higher abundance of myeloid cells as a proportion of total immune cells in IDH-mutant tumors. Our research delves deeply into the diverse cell types and their functional states across both IDH-wildtype and IDH1R132H tumor tissues.

(continue to page 32)

(continued from page 31)

Our findings offer a thorough comparative examination of the immune microenvironment within glioma tissues, shedding light on their functional characteristics depending on their IDH1 mutation status. Our approach has provided a comprehensive characterization of the tumor microenvironments' phenotypic components as well as their spatial neighborhood relationships in human glioblastoma.

Comparison of sample preparation methodologies for cfDNA-based fragmentomics cancer testing

Jason Mitchell[1], Timothy McDaniel[1], Jay Vowles[1], Shannon Davis[1], Denise Butler[1], Jacob Carey[1], Joe Catallini[1], Sian Jones1, Bryan Chesnick[1], Michael Rongione[1], Nicholas Dracopoli[1], Rob Sharpf[1], and Victor Velculescu[1]

1 Delfi Diagnostics, Baltimore, MD, USA

We describe an evaluation of commercially available DNA extraction and library preparation kits conducted as part of the development of a fragmentomics-based cfDNA cancer test. The assessment used a standard set of samples and standardized weighted matrices based on desired performance characteristics. Our laboratory procedure consists of extracting DNA from blood plasma and then applying a standard protocol of end-repair/ A-tailing/ adapter-ligation/ PCR, generating libraries that are shotgun sequenced at low depth on the Illumina NovaSeq 6000 system. To select a DNA extraction kit, we evaluated six different extraction options: Qiagen Circulating Nucleic Acid, PerkinElmer Chemagic cfNA 5k, Beckman Coulter Apostle MiniMax, ThermoFisher MagMax cfDNA isolation, ThermoFisher MagMax cfTNA isolation, and Active Motif Cell Free DNA Purification kits. Each kit was evaluated using three replicates of seven unique plasma samples and scored based on a matrix encompassing yield, fragment size variation, variation of a laboratory-specific fragmentomic score, cost, hands-on time, turnaround time, and operator-assessed usability. To select the library preparation kit, we considered seven different options: NEBNext, NEBNext Ultra II, PerkinElmer BIOO NextFlex, Roche KAPA HyperPrep, ThermoFisher Collibri, Twist Library Preparation Mechanical Fragmentation, and Qiagen QIAseq cfDNA kits. Each option was evaluated using three replicates of six unique cfDNA samples and scored based on a matrix encompassing library yield, percent adapter dimer, conversion efficiency, repeatability and accuracy of a laboratory-specific fragmentomic score, alignment rate, GC content, cost per sample, hands-on time, turnaround time, and operator-assessed usability. Based on our scoring matrices we selected ThermoFisher MagMax cfDNA extraction for DNA isolation and Roche KAPA HyperPrep kit for library preparation. We note that the scoring matrices were optimized for our specific application and we could have drawn different conclusions for other applications. We also note that some commercial options that are available today were not assessed because they were not available at the time of evaluation. Despite these limitations, we believe the approach described here could be generally useful for selecting among NGS sample preparation options.

Comprehensive spatial analysis of metastatic prostate cancer: Tracing metastatic tumor subclones and unveiling microenvironment dynamics

Mengxiao He[1], Sandy Figiel[2], Wencheng Yin[2], Renuka Teague[2], Joakim Lundeberg[1] and Alastair D Lamb[2]

Department of Gene Technology, KTH Royal Institute of Technology, Science for Life Laboratory, Solna, Sweden.
Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK.

The progression of prostate cancer to metastatic disease remains a critical challenge in clinical oncology. Building upon our previous study that explored genomic variations spatially in benign and malignant prostate tissue by inferring copy number alterations (CNA) from spatial transcriptomics data, our follow-up study delves into the intricate landscape of metastatic prostate cancer. In order to explore the dynamics of metastasis and its microenvironment we perform spatial transcriptomics on the entire prostate axial disk, as well as patient-matched lymph node metastases from seven individuals who underwent radical prostatectomy to further shed light into what facilitates certain cancer subclones to invade nearby tissue and disseminate to other sites as well as what sets them apart from other cancer subclones.

We obtained transcriptional information from over 1,000,000 barcoded regions of organ-wide prostate and lymph node metastases and infer genome-wide CNA from these spatially resolved mRNA profiles in situ. In line with previous findings, our data reaffirms the heterogeneous nature of prostate cancer with the presence of many distinct tumor subclones, each characterized by unique genetic signatures. Consistent with previous results, we also find distinct subclones from the primary tumor in different metastases as well as within individual metastases. We constructed a clone-tree to describe sequential clonal events versus independently arising cancer subclone which allows us to spatially trace the evolutionary trajectories of these cancer subclones. Furthermore, by leveraging the spatial information, we gain the unique possibility to explore the microenvironment dynamics of metastasizing tumor subclones. Interactions between cancer cells and the microenvironment is known to facilitate the invasion of nearby tissues and the dissemination of cancer cells, which further underscores the multifaceted nature of metastatic progression. And indeed, apart from having a distinct genetic profile, our results also show a microenvironment distinct from other cancer subclones surrounding the metastasizing tumor subclones.

Collectively, our findings not only advance our comprehension of metastatic prostate cancer but also offers valuable insights into the broader context of cancer evolution and disease progression. These insights hold substantial promise for the development of innovative targeted therapies and personalized treatment strategies in the fight against metastatic prostate cancer.

Copy number variant detection without a panel of normals using anchored multiplex PCR and next generation sequencing

Taylor R. Patterson[1], Allison Hadjis[1], Ryan Rogge[1], Mark Rogers[1], Laura Johnson[1], Devin Tauber[1]

1 Integrated DNA Technologies, Inc., Boulder, CO, USA

Copy number variants (CNVs) such as copy gains or losses may be a cause of disease, a symptom, or both. CNVs affecting oncogenes or tumor suppressor genes are one mechanism by which cancers may arise, proliferate, or persist. CNV methods require a baseline to compare against, usually matched normal tissue or a panel of normal samples. Some next-generation sequencing (NGS) methods allow internal, self-normalization methods, but recommend this only for whole genome sequencing data. Here, we introduce a new CNV method which relies only on data from the sample of interest to determine copy number and breakpoints that is designed to work for targeted Anchored Multiplex PCR (AMP™) panels (RUO). Our prototype method corrects primer read counts for GC%, PCR, and sequencing biases after deduplication using molecular barcodes incorporated during AMP library preparation. Outliers are removed, then the remaining bias-corrected values are segmented. The baseline is estimated from bias-corrected counts of autosomal primers, then used to calculate fold changes and to determine the significance of differences between segments and the baseline. In development testing of more than 200 formalin-fixed, paraffin-embedded (FFPE) tissue and cell line reference inputs prepared with panels ranging in size from 499 to 2595 primers, this prototype method can detect single copy gains and losses, homozygous deletions, and aneuploidy, with specific resolution dependent on a panel's primer distribution. In addition to sensitive detection, our prototype method estimates fold change values highly concordant with ddPCR fold changes. This prototype AMP-based NGS CNV method demonstrates strong performance without requiring paired normal tissue or a panel of normals. In future work, we intend to incorporate the CNV outputs of this method in further calculations such as allele-specific copy number (ASCN) and HRD status determinations.

Elucidating the diversity of tumor microenvironment and EGFR vIII splicing variants in spatial context via in situ sequencing in glioblastoma

Vaibhav Jain[1], Lauren Whaley[1], Diane Satterfield[2], Elizabeth Thomas[2], Emily Hocke[1], Alan Smith[1], Karen Abramson[1], Giselle Lopez[2], Kyle Walsh[5], Michael Brown[3], David M Ashley[3], Roger McLendon[4], Simon Gregory[1]

1 The Molecular Genomics Core, Duke Molecular Physiology Institute at Duke University, Durham, NC, United States.
2 The Preston Robert Tisch Brain Tumor Center at Duke University, Durham, NC, United States.
3 Neuro Oncology, Duke University, Durham, NC, United States
4 Duke Cancer Center Brain Tumor Clinic, Durham, NC, United States
5 Neurosurgery, Duke University, Durham, NC, United States

Glioblastoma (GBM) is a highly aggressive primary brain tumor with a median survival of 12.1 months. This high-grade glioma remains refractory to current treatments like surgery, radiotherapy, and chemotherapy that only extend survival by 2.5 months. GBM's diverse cellular environment necessitates innovative approaches to characterize their heterogeneity and determine the expression profiles associated with tumor progression. Here, we report the first use of 10xGenomics' Xenium in situ sequencing (ISS) platform and a panel of human brain and GBM associated genes to characterize primary and recurrent GBMs, including two pairs of matched tumors. The panel also included a field-first prognostic splicing assay.

Platform based and open-source analysis revealed a high-resolution spatial perspective of the heterogeneous GBM microenvironment. We were able to clearly distinguish genes and pathways being differentially regulated between aggregated primary and recurrent GBM, for example cell cycle pathways being upregulated in primary vs recurrent GBM glial cell clusters and a down regulation of ERBB4 signaling in vascular cell clusters in primary GBMs, which may show vasculogenic mimicry and the formation capillary like structures without endothelial cells. Our data also suggests that SOX11, known to be involved in tumorigenesis, are >2-fold upregulated in annotated tumor cells in primary compared to recurrent tumors. Dysregulated splicing mechanisms are a hallmark of glioma, including alterations in the transmembrane receptor tyrosine kinase, Epidermal Growth Factor Receptor (EGFR), that contributes to GBM malignancy. EGFR variant III (EGFRvIII) is a constitutively active EGFR mutant characterized by loss of exons 2-7 and the ~20% of GBM tumors expressing this variant are associated with conventional therapy resistance and uncontrolled cell growth. For the first time, our data has been able to reveal that EGFRvIII expression can found subpopulations of tumor cells characterized by oligodendrocyte-like and oligodendrocyte progenitor cell (OPC)-like gene signatures, in a spatial context.

(continue to page 37)

(continued from page 36)

Our ongoing analysis shows that ISS approaches serve as a powerful tool for characterizing the heterogeneity of GBM. Characterizing this dynamic tumor environment and characterizing the expression of new ISS splicing assays will enhance our understanding of critical cell-cell interactions mediating the complex TME signaling network, potentially yielding new intervention strategies.

Esophageal adenocarcinoma ecotypes mediate chemoradiation resistance

Rui Ye[1,2,4], Yawei Qiao[1], Shanshan Bai[2], Emi Sei[2], Min Hu[2], Jianzhong He[1], Nan Li[1], Yifan Wang[1], Pablo Lopez[5], Wei Zhang[5], Fei Ye[5], Brian Weston[6], Emmanuel Coronel6, William A. Ross6, Phillip S. Ge[6], Manoop Bhutani[1,6], Nicholas E. Navin[2,3,4,7], Steven H. Lin[1,4]

1 Department of Radiation Oncology, UT MD Anderson Cancer Center, Houston, Texas, 77030, USA
2 Department of Systems Biology, UT MD Anderson Cancer Center, Houston, Texas, 77030, USA
3 Department of Genetics, UT MD Anderson Cancer Center, Houston, Texas, 77030, USA
4 MD Anderson UTHealth Graduate School of Biomedical Sciences, Houston, Texas, 77030, USA
5 Department of Radiation Oncology Research, UT MD Anderson Cancer Center, Houston, Texas, USA
6 Department of Gastroenterology Hepatology and Nutrition, UT MD Anderson Cancer Center, Houston, Texas, 77030, USA
7 Department of Bioinformatics, UT MD Anderson Cancer Center, Houston, Texas, 77030, USA

Esophageal adenocarcinoma (EAC) is an aggressive and lethal disease with a 5-year survival rate less than 20%. The current standard of care for locally advanced EAC patients is neoadjuvant chemoradiation (CRT) followed by surgery and the pathological response rate is low in clinic. An unresolved question is how the cancer cells and their tumor microenvironment (TME) mediate CRT resistance in EAC. To address this question, we performed single-cell RNA sequencing on 84 longitudinal samples from a cohort of patients (N=47) containing 20 good responders (GR) and 27 non-responders (NR). In total, we obtained 302,918 single cells which enabled us to delineate the phenotypic heterogeneity landscape in EAC at baseline, and identify reprogramming in both the immune and tumor cells during treatment that lead to CRT resistance. We identified 4 major cell type compartments in the EAC TME: lymphoid cells (T cells, NK cells, B cells, and plasma cells), myeloid cells (macrophages, neutrophils, dendritic cells, and mast cells), stromal cells (fibroblasts and endothelial cells) and epithelial cells. Our data shows that the TME enriched in lymphoid cells, including resident and effector memory CD8 T cells, follicular helper CD4 T cells, and B cells was associated with better CRT responses. Moreover, a significant myeloid cell infiltration was observed across most patients during CRT and the infiltrated macrophages from GR harbored a more pro-inflammation phenotype. In addition, we identified 4 major expression subtypes of the cancer cells. We found that patients with higher tumor proliferation potential and higher zinc metabolism level were strongly associated with better and worse CRT responses, respectively. Furthermore, we identified the up-regulation of gastric lineage signature expression during CRT in cancer cells that was specific to NR. In summary, this study has provided valuable insights into the biology of EAC CRT response, and identified novel predictive biomarkers and actionable targets to enhance CRT efficacy.

Examination of methylome and transcriptome data reveals the effects of DNA methylation on RUNX1 function in inv(16) acute myeloid leukemia

Derek E. Gildea[1], Ana Catarina Menezes[2], Tao Zhen[2], NISC Comparative Sequencing Program[3], Laura Elnitski[4], Tyra G. Wolfsberg[1], Paul P. Liu[2]

1 Bioinformatics and Scientific Programming Core, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD
2 Oncogenesis and Development Section, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD
3 NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD
4 Genomic Functional Analysis Section, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD

Aberrant changes in DNA methylation have been associated with many cancers, including acute myeloid leukemia (AML). Runt-related transcription factor 1 (RUNX1), which is frequently associated with hematological cancers, is an important regulator in hematopoietic development. DNA methylation is known to attenuate RUNX1 binding at target sites. A subtype of AML cases, defined by the inversion inv(16)(p13q22) on chromosome 16, results in the production of the fusion oncoprotein CBFβ-SMMHC, which is known to interact with RUNX1. Considering that RUNX1 function can be modulated by DNA methylation, we aimed to investigate how the DNA methylation landscape in inv(16) AML affects disease development. To do this, we compared the methylomes and transcriptomes of Lin-Sca1-c-Kit+ bone marrow cells obtained from pre-leukemic inv(16) mice to those of controls. Enzymatic methyl sequencing data from these cells revealed 14,183 differentially methylated regions (DMRs, p-adj < 0.05), with 95% of these DMRs showing hypermethylation in pre-leukemic mice. Gene ontology (GO) term enrichment analysis of the genes found in proximity to these DMRs revealed biological processes of relevance to AML, including myeloid cell homeostasis (GO:0002262), erythrocyte differentiation (GO:0030218), and regulation of DNA replication (GO:0006275). RNA-seq data obtained from the same cells revealed 1,019 differentially expressed genes (DEGs, absolute log2[fold change] ⩾ 1, p-adj < 0.05). Included among these DEGs were RUNX1 targets that had both hypermethylated promoters and downregulated gene expression. Findings from this study indicate that DNA methylation can modulate the RUNX1 function of regulating the expression of target genes in inv(16) AML.

Exome, lcWGS-based copy number assessment, complete transcriptome, and surface protein expression from the same individual cell with ResolveOME

Katherine A. Kennedy[1], Tia A. Tate[1], Isai Salas-Gonzalez[1], Durga M. Arvapalli[1], Swetha D. Velivela[1], Jeffrey R. Marks[2], E. Shelley Hwang[2], Jay A.A. West[1], Victor J. Weigman[1], Jon S. Zawistowski[1]

1 BioSkryb Genomics, Durham, NC.
2 Department of Surgery, Duke University Medical Center, Durham, NC.
Department of Bioinformatics, UT MD Anderson Cancer Center, Houston, Texas, 77030, USA

While the breadth of information obtained from whole genome sequencing (WGS) studies is unmatched by the exposure of intergenic regulatory elements in addition to protein-coding sequence, in light of current sequencing costs and informatics challenges there is still a demand across applications and indications to focus in exonic space, particularly when considering druggability of targets. Furthermore, while targeted DNA panels are commonly performed for diverse single-cell applications, there exists a resolution gap between comprehensive WGS studies and small, focused panels that are currently employed. We have therefore united IDT v2 exome hybrid capture downstream of the ResolveOME single-cell multiomic workflow to enable single nucleotide variation (SNV) ascertainment of the entire cohort of protein coding genes, while retaining the multiomic strengths of the method. The ResolveOME genomic amplification attributes of completeness, allelic balance and uniformity that fuel WGS studies are also critical for obtaining a robust, on-target exome enrichment that empowers accurate SNV detection. Utilizing ductal carcinoma in situ (DCIS) / invasive ductal carcinoma (IDC) cells from a cohort of 9 patients, we performed 757 single-cell exome enrichments, maximizing plexity to 40 cells/capture. Prior to hybrid capture, the corresponding single-cell genomic libraries were subjected to low-coverage WGS (lcWGS) to ascertain copy number alterations, revealing DCIS/IDC-prototypical and heterogenous chromosomal lesions. Full-transcript RNA-seq and BioLegend oligo-conjugated antibody-based surface protein expression were jointly assessed from the transcriptomic libraries of ResolveOME, together enabling cell type identification and phenotypic cell state inference with BaseJumper software. The data indicate the successful merging of the ResolveOME multiomic workflow with a single cell exome survey, enabling a four-pronged analysis including accurate assessment of CNV in the context of a complete transcriptome and panel protein assessment--thus providing a more economical, scalable solution for understanding tumoral genomic variation for when high-depth WGS studies are not required.

Extensive data towards genome in a bottle benchmarks for a new tumor/normal pair

Justin M. Zook[1], Gail Rosen[2], Jennifer McDaniel[1], Andrew Liss[3], Justin Wagner[1], Genome in a Bottle Consortium

1 Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD USA
2 Department of Electrical and Computer Engineering, Drexel University, Philadelphia, PA USA
3 Department of Surgery, Massachusetts General Hospital, Boston, MA USA

Here, we describe results from initial efforts towards the first tumor/normal benchmark from the Genome in a Bottle (GIAB) consortium and detail on-going sequencing from a variety of technologies. We created the first broadly-consented tumor cell line from a pancreatic ductal adenocarcinoma with matched normal pancreatic and duodenal tissue. We have initial Illumina, ONT, HiC, and single cell WGS data, and are now performing an array of additional WGS measurements (PacBio HiFi, Bionano, Onso, and Element) from a large homogeneous batch of the tumor cell line and paired normal tissue. These data are being made public as they are generated, with all planned for public dissemination in 2023, and we are forming an open working group within the GIAB consortium to develop a tumor/normal benchmark from these data.

Initial tumor cell line data indicates it is relatively homogeneous and stable, with substantial aneuploidy from translocations that cause large deletions. We found ~17 large inversions and translocations and 16 chromosomes with extensive loss of heterozygosity due to missing >30% of one copy, in addition to a few smaller duplications. Single cell sequencing showed that most of the large deletions are in nearly all or all cells, though a couple are in only some cells, consistent with bulk WGS analyses from two batches of cells. Somatic SNVs in non-repetitive regions fell in 4 primary classes: ~4000 and ~2000 in practically all cells in diploid and haploid regions, respectively, and ~400 each in only some cells in diploid and haploid regions. Given this cell line's relative homogeneity and stability, with most CNVs being deletions, it may be a good candidate for near-complete genome assembly of the dominant tumor clone as well as the normal to understand complex variation. In addition, we will align tumor and normal reads to each haplotype of the normal assembly to characterize somatic variants, including variants in minor clones. Building on the methods GIAB and T2T have developed to polish diploid assemblies of GIAB's normal genomes and characterize mosaic variants, we will develop benchmarks for somatic variants on the normal assembly as well as standard references.

From pixels to clinical insights: ultra-deep spatial proteomic profiling of the head and neck cancer tumor microenvironment

Arutha Kulasinghe

Immunotherapies, in particular immune checkpoint blockade, have led to durable benefit in a number of solid malignancies including head and neck cancer. Approximately 15-30% of patients obtain a durable benefit, whilst some can have hyper-progression and immune-related adverse effects. Identifying those patients likely to benefit from early diagnostic or surgical resection specimens can be powerful in personalising immunotherapy treatment. In addition, identifying patient unlikely to respond, and potential pathways of resistance, can provide an opportunity for combination therapy.

In this study, we developed and tested 103-antibodies targeting tumour and immune phenotyping, metabolic and stress pathways in an ultra-high plex, single-cell resolved, whole slide format for head and neck cancers. Optimisation of the panel was performed on tonsillar tissues and clinical testing and validation on a combination of oropharyngeal and oral cavity cancers from pre-treatment tissue samples from patients on immunotherapy trials. Our study identified unique tumour and immune neighbourhoods composed of specialised immune cell subsets, geo-spatially distinct across the tissue. Moreover, pockets of immunotherapy-sensitive regions were identified where there was increased expression of pro-apoptotic markers BAX, BAK, with higher levels of CD8 T-cell infiltration. In other regions of the tumour, increased levels of GLUT-1, G6PD and invasion markers MMP9 identified immunotherapy resistance regions and areas of the leading edge of the tumour, likely to metastasize.

Taken together, our study highlights the use of high-plex, whole-slide, spatial proteomic profiling to identify cellular neighbourhoods and niches of immunotherapy sensitivity and resistance in an integrated, multi-omic manner. We believe this integrated methodology will provide new biomarkers and pathology interpretable feature identification which can be used for the development of companion diagnostics and combination therapies.

HEK293 genome plasticity revealed using long read RNA and DNA sequencing

Bradley W. Langhorst, Laura Blum, Giron Koetsier, Brittany Sexton, Danielle Fuchs, Evan Janzen, Jonathan C Sanford, Luo Sun, Dora Posfai

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

HEK293 cells are used extensively in cancer research, with more than 4599 articles published between 2017 and 2023 (PubMed search "cancer[MeSH Major Topic] and hek293"). Studies involving transfection to overexpress, knock down, or modify proteins in these cells are common. While HEK293 cells are easy to grow and maintain, Yao-Cheng et. al. identified genomic variability between cell line sources and revealed unexpected details about their developmental lineage (likely adrenal, not kidney) [1]. Genomic or transcriptomic reorganization may complicate interpretation of results from targeted intervention experiments.

We compared HEK293 cells from various commercial sources using long read sequencing technology. Oxford Nanopore RNA profiling was used to compare isoform usage and identify gene fusions. Long DNA libraries were prepared using a high molecular weight isolation method, sequenced on Oxford Nanopore, and assembled to produce long contiguous fragments. Genome fragments generated from long read sequencing support the genomic diversity reported by Yao-Cheng et.al, though the specific variants are distinct from those previously reported. Given the plastic nature of HEK293 cells, we recommend confirmation of genomic and transcriptomic structure of the specific cells in use before and after intervention.

[1] Lin, Yao-Cheng, Morgane Boone, Leander Meuris, Irma Lemmens, Nadine Van Roy, Arne Soete, Joke Reumers, et al. "Genome Dynamics of the Human Embryonic Kidney 293 Lineage in Response to Cell Biology Manipulations." Nature Communications 5, no. 1 (September 3, 2014): 4767. https://doi.org/10.1038/ncomms5767.

High-quality single-nuclei RNA sequencing of Glioblastoma and other clinical samples from FFPE tissue with automated samples preparation

Ronan Chaligne[1], Nate Pereira[3], Catherine Snopkowski[1], Brigita Meškauskaitė[1], Meril Takizawa[1], Ignas Masilionis[1], Kenny Yu[2], Kathleen Luckett[2], Viviane Tabar[2], Karuna Ganesh[2], Nabiha Khan[3], John Bashkin[3], and Stevan Jovanovich[3]

1 Single-cell Analytics Innovation Lab, Memorial Sloan Kettering Cancer Center, NY, USA;
2 Memorial Sloan Kettering Cancer Center, NY, USA;
3 S2 Genomics, Inc.

Advancements in single-cell RNA sequencing (scRNA-seq) technologies have ushered in a new era of insights into gene expression at the cellular level. However, the quantification of gene expression in scRNA-seq is profoundly influenced by variations in sample preparation, biochemistry, and RNA capture methods. The conventional approach involves capturing mRNA molecules using a poly-dT sequence, yet this method falls short in lower quality or formaldehyde-fixed samples, resulting in suboptimal quantification. Given the potential impact single cell analyses of FFPE tissue could have on basic and translational research, we sought to improve current methods by automating sample preparation steps in combination with the recently released Chromium Single Cell Gene Expression Flex workflow.

We have demonstrated the efficacy of this approach by running comparative snRNA-seq analyses of flash frozen and FFPE glioblastoma and other human clinical samples. Our preliminary analysis of matching FFPE and frozen samples has revealed comparable library size distributions between the two approaches, with noteworthy statistics such as a median of 7,500 counts per cell and a median of 3,600 genes per cell for fixed tissue, as exemplified with Glioblastoma samples. These results were obtained by fully automated the deparaffinization, rehydration, and processing of FFPE curls into singulated nuclei using a Singulator prototype developed by S2 Genomics, followed by library preparation with the Chromium Single Cell Gene Expression Flex kit, a game-changing technology that employs short barcoded probes templates to quantify gene expression. This innovation opens the door to scRNA-seq and snRNA-seq analysis even in formaldehyde-fixed samples, including FFPE.

At the AGBT conference, we will present comprehensive data showcasing the preparation and analysis of snRNA-Seq libraries from a diverse array of human and murine FFPE samples. This presentation will provide insights into the impact of being able to assess valuable FFPE clinical samples with snRNA-seq.

High-resolution proximity-guided cytogenomics improves patient stratification and uncovers new variants in leukemia and solid tumors

Ivan Liachko[1], Stephen Eacker[1], Maika Malig[1], Kyle Langford[1], Moutusee Islam[1], Alexander Muratov[1], Brad Nelson[1], Olga Sala-Torra[2], Min Fang[2,3], Shawn Sullivan[1], Cecilia Yeung[2,3], and Jerald Radich[2]

1 Phase Genomics, Inc. Seattle, WA
2 Translational Sciences and Therapeutics Division, Fred Hutchinson Cancer Center, Seattle WA
3 Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA

Cytogenetic methods are a cornerstone of the modern diagnostic workflow for liquid cancers and constitutional genetic disorders. The goal of these methods is to identify structural genomic aberrations that drive the cancer or otherwise act as an indicator of patient risk. Despite their success, any one cytogenetic test is limited either in resolution, ability to provide unbiased survey of the genome, or ability to detect balanced aberrations. Additionally, the cascade of reflex testing is inherently slow and difficult to scale and cannot be applied to the bulk of solid tumor samples preserved as FFPE.

To address these limitations, we have applied proximity ligation sequencing (PLS) to characterize structural variants in AML genomes. PLS is a scalable, short-read based method that captures ultra-long-range genomic sequence information without high-molecular-weight DNA and requires a modest (5-7x) sequence coverage of the genome. Importantly, this method is highly versatile in its sample requirements and has been validated through hundreds of samples of solid and liquid tumors, FFPE tissue, and cultured cells.

Applying this method to a set of AML diagnostic specimens, we find that PLS identifies cytogenetically defined translocations (reciprocal and non-reciprocal) and inversions with specificity and sensitivity >0.95. Its high resolution detected non-canonical variants missed by standard methods as well as structural variants below limits of detection by karyotyping, identifying variants in over half the patients previously determined to have a normal karyotype. This included previously described variants of clinical significance and a recurrent inversion not previously described in AML, as well as reclassifying a significant number of patients. These observations highlight the effectiveness of PLS to provide high-resolution insights into known clinically relevant variants and discover variants missed by traditional methodologies.

Integrated analysis of whole genome, whole transcriptome sequencing, and ATAC-seq in renal cell carcinoma

Tatsuhiro Shibata

Laboratory of Molecular Medicine, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan

Renal cell carcinoma (RCC) comprises several histological types characterized by different genomic and epigenomic aberrations; however, the molecular pathogenesis of each type remains to be elucidated. We performed whole-genome sequencing of 128 Japanese RCC cases of different histology to elucidate the significant somatic alterations and mutagenesis processes. We also performed transcriptomic and epigenomic sequencing to identify distinguishing features, including assay for transposase-accessible chromatin sequencing (ATAC-seq) and methyl sequencing. Genomic analysis revealed that the mutational signature differed between histological types, suggesting that different oncogenic factors drive each histology. Clear cell RCC was classified into three epi-subtypes, one of which expressed high levels of immune checkpoint molecules with frequent loss of chromosome 14q. From the ATAC-seq results, master transcription factors were identified for each histology. In particular, we found a high contribution of GRHL2 in chromophobe RCC. As GRHL2 is involved in distal nephron differentiation, this suggests the cell of origin of this tumor type. By integrating WGS and ATAC-seq, we identified somatic promoter mutations that increase chromatin accessibility. These genomic and epigenomic features may lead to the development of effective therapeutic strategies for RCC.

Isabl GxT: Rapid and scalable whole-genome and transcriptome sequencing workflow for oncology

Max F. Levine,1 Aditya S. Deshpande,1 Kevin Hadi,1 Minal Patel,1 Stan Skrzypczak,1 Juan S. Medina-Martínez.1

1 Isabl Inc., New York, NY

Cancer whole-genome and transcriptome sequencing (cWGTS) provides a comprehensive view of a tumor's genome, spanning mutation classes, clonal phylogenesis, and mutation signatures. While cWGTS is becoming increasingly accessible, data analysis and interpretation remains complex, necessitating specialized expertise and robust computational infrastructures.

Isabl has developed scalable, fast, and cost-effective sample-to-report and informatics solutions spanning applications in oncology, including analysis of fresh frozen and FFPE biospecimens from solid tumors, and hematological neoplasms.

Isabl GxT, an FDA Breakthrough Designated integrated assay (90x Tumor, 40x normal germline control and 60M paired-end RNA-seq), for solid tumors has demonstrated comparable sensitivity to high-depth, FDA-approved clinical diagnostic assays. From FASTQ to report, results are generated in less than 15 hours.

By conducting ensemble variant calling, the assay performs highly confident identification of germline and somatic mutations, allele-specific detection of aberrant copy number segments, with integration to gene expression. We developed a proprietary pan-cancer Homologous Recombination Deficiency (HRD) classifier (Isabl HRD) which leverages diverse mutational classes and signatures to yield highly sensitive and accurate classification scores comparable to HRDetect and CHORD. Isabl GxT analytics processed data from 413 (90x) paired tumor normal genomes from FASTQ to report in 5 days, resulting in greater than 60 million passed variants including 2,187 drivers.

(continue to page 48)

(continued from page 47)

Isabl GxT Heme, a pan-heme (myeloid, lymphoid, and plasma cell) assay and analytical workflow addresses the unmet need in hematological malignancies. Isabl GxT Heme leverages Isabl Heme DB, our proprietary variant database annotating 40% more variants missed by established annotation databases that are primarily informed by solid tumor data. Compared to standard of care assays, Isabl GxT Heme showed 100% sensitivity in detecting recurrent SVs (41/41) and CNVs (67/67) against cytogenetics. For recurrent SNVs (219/236) and indels (109/117), its sensitivity was 93% when compared to panel sequencing.

Isabl GxT provides an interactive web navigator supporting single sample and project level reports, dynamic queries, and download of publication-ready figures.

The Isabl platform provides an unparalleled all-in-one solution for cancer whole genome and transcriptome analysis driving discoveries in cancer research and treatment for both solid and hematological neoplasms.

Matched fresh frozen and FFPE patient tissues reveal the enhanced sensitivity and data quality of a novel DNA library prep method

Margaret R. Heider, Jian Sun, Brittany S. Sexton, Laura Blum, Bradley W. Langhorst, and Pingfang Liu

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

FFPE DNA poses many notable challenges for preparing NGS libraries, including low input amounts and highly variable damage from fixation, storage, and extraction methods. It is difficult to obtain libraries with sufficient coverage and the sequencing artifacts arising from damaged DNA bases confound somatic variant detection. Additionally, many laboratories process FFPE tumor samples alongside matched, high quality, normal DNA and many library prep workflows are not readily compatible with both sample types.

We developed a novel NGS library prep method compatible with both high quality and very low quality FFPE DNA samples, employing three new enzyme mixes designed specifically for compatibility with FFPE samples including a DNA repair mix, an enzymatic fragmentation mix, and a high yield PCR master mix. To validate this workflow on real patient samples, we obtained DNA extracted from matched tumor and normal tissue of various tissue types preserved by both fresh frozen and FFPE methods, with fresh frozen-extracted DNA providing the gold standard for library quality and mutation content. The FFPE DNA samples ranged in quality from DNA integrity number (DIN) 1.5 to 6.8. We prepared libraries using this method compared against other library prep workflows and sequenced them by WGS and target capture.

This new enzymatic fragmentation-based library prep workflow not only reduced the false positive rate in somatic variant detection by repairing damage-derived mutations in FFPE DNA samples but also improved the library yield, library quality metrics (including mapping, chimeras, and properly-paired reads), library complexity, coverage depth and uniformity, and hybrid capture library quality metrics. Comparing the variant calls from matched FFPE and frozen tissues revealed an improved sensitivity and accuracy of variant calling using this library prep method compared to mechanical shearing and other enzymatic fragmentation library prep approaches.

This new suite of enzyme mixes improves the overall library prep success rate from challenging FFPE samples, allowing even highly damaged FFPE samples to achieve high quality libraries with a greater sensitivity for somatic variant identification. The workflow is robust and flexible, compatible with both FFPE DNA and matched high quality DNA samples as well as automation-friendly for convenience in sample processing.

On-boarding and standardization of Enzymatic Methyl Sequencing to provide accessible high-throughput, high-resolution whole genome methylome analysis for oncogenomic research.

Kristie Ng[1], Andrea Bevan[1], Lubaina Kothari[1], Savo Lazic[1], Madhuran Thiagarajah[1], Ting Liu[2], Mohamed Alias[2], Michael Hong[2], Paul Wambo[3], Robert Kridel[2], Shraddha Pai[3], Bernard Lam[1]

1 Translational Genomics Laboratory, Ontario Institute for Cancer Research, Toronto, ON, Canada
2 Division of Medical Oncology & Hematology, Princess Margaret Cancer Centre, Toronto, Ontario, Canada.
3 Adaptive Oncology, Ontario Institute for Cancer Research, Toronto, Canada

DNA methylation has dynamic influences on cancer initiation and progression and can be an invaluable biomarker for early cancer diagnosis and treatment. Current methylome analysis assays include Methylated DNA Immunoprecipitation Sequencing (MeDIP-seq) and Whole Genome Bisulfite Sequencing (WGBS), both of which have benefits and disadvantages. MeDIP-seq offers enrichment of methylated sequences while maintaining DNA integrity, however these assays lack resolution. WGBS provides single-nucleotide resolution, but chemical treatment degrades DNA and causes GC content bias. Enzymatic Methyl-seq (EM-seq) uses enzymatic conversion to provide identification of 5-methylcytosine and 5-hydromethylcytosine with single nucleotide resolution and minimal DNA degradation. Combining EM-seq with cell-free DNA samples provides a non-invasive method for tumour DNA methylome profiling with promising applications for early cancer detection, assessing tumour burden, and monitoring minimal residual disease.

The Translational Genomics Laboratory, part of the Ontario Institute for Cancer research (OICR), a Canadian translational cancer research centre, is dedicated to leading the way in clinical cancer care by providing cutting edge advancements in Next Generation Sequencing. Our clinical arm is committed to bringing high-quality assays into compliance with CAP/ACD-accredited and CLIA certified standards to deliver testing and analysis to clinicians. Our R&D arm focuses on onboarding oncogenomic assays, standardizing and scaling protocols for high-throughput production, and integrating optimized data analysis through our custom informatics pipeline.

(continue to page 51)

(continued from page 50)

In this presentation, we demonstrate our process of onboarding EM-seq as a routine offering for research uses. We have demonstrated that the assay generates high-quality libraries with sample inputs as low as 10 ng of solid tumour DNA and 1 ng of cell-free DNA. We observe an average conversion rate of 96.4% for CpG methylated pUC19 DNA and 0.4% for unmethylated Lambda DNA, demonstrating successful enzymatic conversion for solid tumour and cell-free DNA samples. Our EM-seq offering has been used by collaborators studying epigenetic biomarkers of early progression in glioblastoma, altered DNA methylation in medulloblastoma, and methylome signatures of circulating lymphoma tumor DNA for non-invasive assessment of treatment response. Next, we plan to increase sample throughput using automated liquid handling and to bring this assay into compliance with CAP/ACD-accredited, CLIA-certified, ISO 15189 standards for use in clinical projects.

Optimized biospecimen procurement in support of personalized oncogenomics

Andrew Mungall[1], Robin Coope[1], Natasja Wye[1], Rebecca Carlsen[1], Pawan Pandoh[1], Carrie Hirst[1], Jacqueline Schein[1], Jessica Nelson[1], Yongjun Zhao[1], Janessa Laskin[2], Marco A. Marra[1,3,4]

1 Canada's Michael Smith Genome Sciences Centre, BC Cancer, Vancouver, British Columbia, Canada.
2 Department of Medical Oncology, BC Cancer, Vancouver, British Columbia, Canada.
3 Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada
4 Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia, Canada

The application of "omics" technologies to precision oncology relies on biospecimen procurement and preservation protocols designed to limit the effects of pre-analytical variables. These include cold ischemia time (defined as the amount of time that the sample is at ambient temperature following tissue removal), rate and method of cryopreservation or fixation, sample storage time and temperature, and the number of freeze-thaw cycles. Rapid processing of tissue post-biopsy is crucial to preserve the physiological state of cancer cells, which is necessary to accurately identify somatic alterations in the tumour genome, transcriptome (including gene expression) and epigenome, and to delineate intra-tumour heterogeneity.

Sample freezing rates were compared between various methods of biospecimen tissue freezing and resulted in the design and manufacture of custom aluminum trays for holding Cryomolds® on dry ice. A robust procedure was developed for optimal preservation of needle core and other small surgically excised tissue biopsies embedded in optimal cutting temperature compound within the biopsy suite. Educational materials, including videos, were also developed to facilitate the training of tissue procurement staff at remote cancer centres and increase the rate of collection of cryopreserved tumour specimens.

We present the protocol used for the successful procurement and preservation of biopsies from over 1,500 advanced cancer patients and subsequent whole genome and transcriptome analyses, resulting in the production of high quality data used to inform cancer treatment planning.

Presence of DNA-like biotypes in cfRNA libraries persist after standard DNase treatment and should be considered when investigating cfRNA biomarkers.

Anna Hartwig, Jeremy Ku, Magdalena Gebala, Seda Kilinc, Alice Huang, Dang Nguyen, Diana Corti, Lisa Fish, Babak Alipanahi

Background: Exai Bio develops cell free (cf)RNA-based tests, using artificial intelligence to identify patterns in a subset of cancer-specific small RNAs called orphan non-coding (onc)RNAs. Profiling cfRNA biotypes is foundational for discovery and utilization of cfRNA-derived biomarkers. Notably, while cellular RNA is largely made up of rRNA, biotypes from cfRNA-seq data vary between studies, and often show sizeable levels of DNA repeat elements (reviewed in Turchinovich et al, Front. Immunol, 2019). Here, we investigate cfRNA biotypes from different blood tubes and use nuclease treatments to elucidate their origin.

Methods: Samples from donors with or without lung cancer (n=192) were drawn into serum, EDTA plasma, and Streck plasma tubes. The resulting biofluids underwent bead-based extraction including on-bead DNase I treatment, small RNA library prep and sequencing. Abundance of biotypes was calculated, and further investigated by nuclease treatments of RNA from separate samples.

Results: All biofluids tested generated acceptable RNA yields, library conversion, and sequence metrics. Serum samples generated higher average nucleic acid extraction yields than plasma. (13.3 ng/ml biofluid vs 2.5 and 2.7 ng/ml for EDTA and Streck, respectively.) The top biotypes detected were repeats: SINEs in serum (17% of read annotations), and segmental duplication (segdup) in EDTA (18.1%) and Streck (17.7%). We hypothesized that some of these repeats could be derived from DNA contamination of the RNA library. While the extraction method used contains on-bead DNase I treatment, this treatment may be incomplete, and less effective for degrading ssDNA. RNase and ssDNase treatments identified biotypes that are sensitive to ssDNase treatment and insensitive to RNase treatment (DNA-like) and vice-versa (RNA-like). Examples of DNA-like biotypes include LINEs and SINEs. Examples of RNA-like biotypes include miRNAs, snoRNAs and rRNAs. Interestingly, “other repeats” and “segdup” were more RNA-like, suggesting that these represent transcribed elements. Finally, we demonstrate that oncRNAs currently used as features in Exai Bio’s cancer models are RNA-like.

Conclusion: Our study underlines the importance of characterizing the biotypes of interest in cfRNA data and to be aware of DNA as a contributor in cfRNA libraries, even after standard DNase I treatment.

Sensitive and accurate detection of small RNAs in tumor and normal tissues.

Deyra N Rodríguez, Heather Raimer Young, Brittany S Sexton, Kaylinnette Pinet, Gautam Naishadham, Brad Langhorst, Louise Williams

New England Biolabs, Ipswich, MA 01938, USA

Small noncoding RNAs (sncRNAs) play fundamental roles in cancer development, progression, and drug resistance. Increasing evidence shows that sncRNA expression signatures can be used to discriminate between normal and cancer tissues, highlighting their potential as diagnostic and prognostic biomarkers of the disease. Next generation sequencing (NGS) is a powerful tool for the comprehensive detection and sequence characterization of sncRNAs. We have developed an enhanced method to detect sncRNA that has increased sensitivity and accuracy when compared to current commercially available kits. This method reduces the ligation bias present in ligation-based library preparation workflows.

Here we demonstrate the power of this method to investigate the expression signatures of sncRNAs in 6 different human tissues (brain, testis, placenta, bladder, ovary and esophagus) as well as 5 sets of matched normal and tumor samples (breast, stomach, rectum, colon, and liver). We found that sncRNAs, included micro RNAs (miRNAs), piwi-interacting RNAs (piRNAs), small nucleolar RNA (snoRNAs), tRNA-derived small RNAs (tsRNA) demonstrate tissue specific expression; with human brain and placenta expressing the higher number of miRNAs. Analysis of tumor and normal tissues identified differentially expressed sncRNAs within a match pair as well as between tissue types. This low bias small RNA library preparation method increased our confidence in the small RNAs detected. This method is robust and can be used for studies investigating biomarker selection.

Sensitive, distributable, and cost-effective molecular residual disease testing using whole genome cell free DNA sequencing

James Han, Kyria Roessler, Sven Bilke, Fan Song, Yunjiao Zhu, Jeff Tsai, Li Liu, Khai Luong, Emily Parker

Molecular residual disease (MRD) detection using circulating-tumor DNA (ctDNA) is increasingly utilized in cancer treatment for monitoring therapy response, relapse detection, and as a surrogate endpoint in clinical trials. A variety of technical approaches are being employed by laboratories. Tumor-informed MRD assays are based on a patient-specific fingerprint of tumor associated variants (SNV, indels, CNV, SV, methylation, etc.) which are later detected from cell-free DNA (cfDNA) using a patient bespoke panel, a fixed panel, or whole genome sequencing (WGS). Tumor naïve MRD assays locate tumor specific biomarkers in cfDNA without a tumor fingerprint. Tumor informed assays are generally more sensitive than tumor naïve assays. While tumor naïve assays have lower sensitivity, they have the potential to identify molecular biomarkers when tumor DNA is unobtainable.

The majority of commercially available MRD assays sequence cfDNA with high depth to detect the presence of tumor-specific somatic variants previously identified from a patient's tumor tissue or baseline pre-surgery plasma. Amongst the tumor informed approaches, tracking tumor mutations in cfDNA with patient bespoke probes has become a popular approach due to its potential for high analytical accuracy (down to 100 ppm) and low overall sequencing cost. However, patient bespoke approaches require high DNA input for high accuracy, longer turnaround times, and operational complexity when running a patient-specific assay. Fixed panels have the potential for high accuracy. Creating a panel with high sensitivity across many tumor types can be a challenge due to tumor type specific mutational rates and patterns.

We have developed a simple, automated, distributable, and cost-effective tumor-informed pan cancer WGS workflow that can detect ctDNA down to 10 ppm. Using WGS for both the tumor tissue and matched normal, the approach can yield a broad, patient and tumor-specific set of somatic variants for MRD detection. With subsequent low-pass WGS of the patient's plasma sample combined with a statistically driven, breadthwise analysis approach, the entirety of the unique fingerprint of tumor somatic variants can be monitored. We demonstrate that this technique can work across a range of tumor mutational burden (TMB), achieving detection to 10-100ppm using 5ng of cfDNA input from 2-4mL of plasma.

Single cell spatial transcriptome analysis of LCNEC (Lung Cancer) using STOmics

Akinori Kanai[1], Ryota Matsuoka[2], Daisuke Matsubara[2], Ayako Suzuki[1], Yutaka Suzuki[1]

1 Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo
2 Department of Diagnostic Pathology, Institute of Medicine, University of Tsukuba

A lot of single cell analysis has been performed because of the low sequencing costs and the improvement of the analysis methods. 10,000 cells can be analysed in single cell analysis using 10x genomics NEXT GEM technology. However, spatial location information is lost when tissue is dissociated into single cell. We don't know where each cell was located, which genes were expressed and what roles they played.
We performed STOmics which can analyze gene expression at the single cell resolution having spatial positional information. The sections of fresh frozen tissue are placed on a chip with oligos poly-T sequences spaced 500 nm apart in the center of the spot. The mRNA is captured on the chip to create a library for NGS. MGI T7 or G400 sequencers were used for data acquisition.

First, STOmics analysis was performed using mouse brain, and it was confirmed that clustering at the bin 200 was consisitent with the histological image, and that the gene expressions were separated by histological region. We also confirmed that the clustering at the cell-bin showed gene expression that matched the tissue image. Ttr high expressing cell bin cluster was located in the choroid plexus.
Next, STOmics was applied to large cell neuroendocrine carcinoma (LCNEC) specimens for further analysis. The distribution of cell markers and small cell carcinoma (SCLC) subtype markers were consistent with data of STOmics and Visium. The Neuroendocrine (NE) tumor cells were divided into multiple clustered. The NE tumor cells were heterogeneous, and one of them was characterized by cancer stem cell marker gene. This cancer stem cell cluster may have been the source of other NE tumor cells. We would like to present the new findings obtained from the single cell spatial analysis.

Spatial omics analyses reveal molecular relationships between tumor cell progression and distinctive microenvironment in lung adenocarcinoma tissues at a single-cell resolution

Ayako Suzuki[1], Yuma Takano[1], Jun Suzuki[2,3], Kotaro Nomura[2], Genichiro Ishii[4], Masahiro Tsuboi[2], Yutaka Suzuki[1]

1 Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan
2 Department of Thoracic Surgery, National Cancer Center Hospital East, Chiba, Japan.
3 Department of General Thoracic Surgery, Juntendo University Hospital, Tokyo, Japan
4 Department of Pathology and Clinical Laboratories, National Cancer Center Hospital East, Chiba, Japan.

To understand molecular mechanisms of lung cancer progression, we performed spatial omics analyses for eight lung adenocarcinoma cases. Using spatial transcriptome sequencing Visium, we first defined transcriptome trajectories of tumor cell progression and extracted distinctive microenvironment features, including immune cells and stromal cells located in the boundary regions of transcriptomic and/or histological changes of tumor cells. We then validated the identified molecular events and relationships between tumor cells and their surrounding cells at a single-cell level by conducting multiplexed immunofluorescence PhenoCycler and in situ gene expression profiling Xenium.

As a result, tumor cell progression from well-differentiated to more malignant proliferative/invasive regions were associated with immune cell elimination and anti-inflammatory macrophage infiltration. In Xenium data, we verified that SPP1+ anti-inflammatory macrophages were enriched in proliferative/invasive regions at a single-cell resolution.

We further analyzed another adenocarcinoma case with heterogeneous histological features. A distinct transcriptome cluster was identified in the boundary region between well-differentiated and moderate to poor/mucinous regions. Genes associated with immune response were highly expressed in this region. For detailed analysis of this region at a single-cell level, we evaluated the PhenoCycler and Xenium data and found that there were cytotoxic CD8+ T cells with granzyme and perforin expression while IDO1-expressing tumor cells and CCL22+ cells, which were known to be associated with immune suppression, uniquely located in this region. The results would indicate that there is confliction between immune active and suppressive players in boundary regions of transcriptomically/histologically different tumor cell clusters. Spatial omics analyses enable identifying such key regions and key molecules for understanding and prevention of lung cancer progression.

The TET-OGT interaction maintains DNA methylation and suppresses expression of transposable elements in mouse embryonic stem cells

Hugo Sepulveda*, Xiang Li*, Carlos Angel, Leo Arteaga, Caitlin Brown, Melina Brunelli, Patrick Kennedy, Cindy Manriquez, Xiaojing Yue, Robert Crawford, Fabio Puddu, Jamie, Scotcher, Walraj Gosal, Jens Fullgrabe, Nikolay Pchelintsev, Páidí Creed, Samuel A Myers, Geoffrey J. Faulkner and Anjana Rao

The O-GlcNAc transferase OGT binds TET methylcytosine dioxygenases through a conserved sequence located at the C-terminal end of all three mammalian TET enzymes. We show here that deletion of the Ogt gene in mouse embryonic stem cells (mESC) results in a global loss of the TET substrate 5-methylcytosine (5mC) in both euchromatic and heterochromatic compartments, with a concomitant increase in the TET product 5-hydroxymethylcytosine (5hmC) at the same genomic regions. Discrimination of 5mC and 5hmC was enabled by 'six-letter sequencing', an enzymatic single-workflow technology for the simultaneous detection of both cytosine modifications (alongside canonical bases G, C, T and A) at base-resolution and from a single short-read sequencing run (in development by biomodal).

The widespread change in DNA cytosine modification patterns is accompanied by de-repression of transposable elements (TEs) that bear the O-GlcNAc modification and are normally tightly suppressed in heterochromatin. mESC engineered to lack the TET1-OGT interaction resemble OGT-deficient mESC in displaying a global loss of 5mC. We conclude that TET proteins, particularly TET1, have two opposing biological functions – their canonical enzymatic activity results in loss of DNA methylation through oxidation of 5mC, but the TET1-OGT interaction maintains DNA methylation across the mammalian genome and silences LINE-1 and LTR expression in heterochromatic regions of the genome.

Third gen methods for characterization of an osteosarcoma cell line

DW Mohr[1], AF Scott[1], WA Littrell[1], M Kokosinski[2], YS Zou[3], V Stinnet[3], J Madiwala[4], J Anderson[4], K Gabrielson[4,5]

1 Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan
2 Department of Thoracic Surgery, National Cancer Center Hospital East, Chiba, Japan.
3 Department of General Thoracic Surgery, Juntendo University Hospital, Tokyo, Japan
4 Department of Pathology and Clinical Laboratories, National Cancer Center Hospital East, Chiba, Japan.

We have characterized the widely used rat osteosarcoma cell line UMR-106 using a variety of techniques including Oxford Nanopore (ONT) ultra-long reads and 5mC identification, Bionano Genomics (BNG) optical maps, and Illumina (ILM) short-read RNAseq. The N50 for the ultralong reads was 75kb. The top 20 candidate genes from human studies of osteosarcoma reported in the Catalogue of Somatic Mutations in Cancer database (cancer.sanger.ac.uk/cosmic) were inspected using IGV as well as KRAS, CDK1, IFITM5, MYC, NOTCH1, BRCA2 and PRKAR1A. Likely pathogenic mutations from COSMIC were found in H3F3A, TP53, BRAF, KMT2D, and MYC. SVs in the epigenetic marker genes H3F3A and KMT2D implied that methylation abnormalities might occur in UMR-106 and led us to look for 5mC using nanopore sequencing by comparing the cell line to standard ONT reads (~15kb N50) from a Sprague-Dawley control animal, the strain from which the cell line was derived in the 1970s. Significant differences in methylation patterns were seen between cell line and normal control which may reflect, in part, differences in tissue type (tumor vs lymphocyte) as well as oncogenicity. Optical mapping identified 9,182 deletions, 2,3818 insertions, 130 duplications, 418 inversion breakpoints, 249 interchromosomal translocation breakpoints, and 107 intrachromosomal translocation breakpoints when aligned to DLE1 sites.

We compared expression results to publicly available rat osteoblast data and noted overexpression of genes that showed genomic amplification, including MYC which was within an amplified block of several Mb. Both the amplification and the presence of a "human" activating c-MYC mutation are consistent with it being a precipitating event in the origin of the UMR-106 tumorigenicity. Further, the similarity between oncogenic mutations in both UMR-106 and those reported in patients suggest that this cell line is a good model for osteosarcoma. The combination of methods that we used illustrates the power of newer technologies to rapidly characterize cell lines and we argue that such genomic analyses should be done for all cell lines widely used so that experimental interpretation of results can be better understood.

TopoLink: A rapid, high-resolution proximity ligation method for the detection of structural variants and chromatin topology features in cancer.

Meredith Carpenter[1], Jonathon Torchia[1], Mital Bhakta[1], Philip Uren[1], Lisa Munding[1]

Cantata Bio, LLC, Scott's Valley, California, USA.

Despite numerous technological advances, including the widespread adoption of next-generation sequencing, many clinically relevant cancer driver mutations go undetected. The reason for this gap in understanding is two-fold: 1) Structural variants (SVs) are particularly challenging to detect using shotgun or long read-based approaches, particularly in heterogenous samples; and 2) It is increasingly accepted that epigenetic dysregulation may be a driver of disease state through mechanisms such as enhancer hijacking or changes to 3D genome organization.

To address this gap in understanding, we developed TopoLink: a dual-purpose proximity ligation library protocol that yields high-quality, high-resolution, unbiased Hi-C libraries capable of sensitive detection of structural variants and chromatin topological features from a single data set. Importantly, the TopoLink protocol can be performed in under 8 hours – less than half of the time required for conventional Hi-C protocols.

In the current study, we demonstrate the performance of TopoLink for the detection of variants within the primary sequence (such as SVs, SNVs/Indels, and CNVs) in addition to high-resolution reporting of 3D chromatin conformation. We performed benchmarking analyses for both genetic variant detection and chromatin topology in two well-characterized lymphoblast cell lines (GM12878 and K562). We demonstrate the capability of TopoLink data to detect interchromosomal translocations with 10-fold higher sensitivity over shotgun. Furthermore, using hybridization capture to detect the BCR-ABL1 fusion in K562 cells, we show that fewer than 100k read pairs is sufficient to detect this fusion with high confidence in TopoLink libraries. We find that the shotgun-like, uniform coverage of TopoLink data also enables detection of SNVs/Indels with improved sensitivity and precision over conventional Hi-C methods. Importantly, we demonstrate that TopoLink libraries maintain topological features consistent with biologically relevant topologically-associated domains (TADs) and chromatin loops, providing insight into novel epigenetic cancer drivers. TopoLink enables higher-resolution assessments of chromatin topology over conventional multi-RE-based Hi-C approaches, as demonstrated by 50% more TADs and loops called at 5-kb resolution. Taken together, our data demonstrate that TopoLink is a next-generation Hi-C method that enables sensitive detection of clinically relevant structural variants and high-resolution chromatin topology in a single, <8-hour workflow.

Identifying cancer neighborhoods and cancer-associated fibroblasts in prostate cancer using MERFISH

Ben Patterson[1], Cheng-Yi Chen[1], Nicolas Fernandez[1], Simran Kaushal[1], Leiam Colbert[1], Jiang He[1]

1 Vizgen, Cambridge, MA, USA

Prostate cancer is highly heterogeneous due to the presence of competing clonal expansions with independent origins. Preserving the spatial context of prostate cancer tissue is critical when studying the gene expression profiles across heterogeneous cancer cell populations. Built on Multiplexed Error-Robust Fluorescence in situ Hybridization (MERFISH) technology, the Vizgen MERSCOPE® Platform provides a robust method to study biological tissue in this manner by directly profiling of the spatial organization of intact tissue with subcellular resolution. Here, we used a 500-gene Predesigned Immuno-oncology Panel to assess the canonical signaling pathways of cancer, cancer type-specific genes, select immune genes, proto-oncogenes, and tumor-suppressor genes in patient-derived prostate cancer samples. We directly imaged the expression of marker genes for cancer cells, fibroblasts, immune cells, and endothelial cells in these malignant tumors and clustered the cell types and states based on their gene expression profile. Through this analysis, we identified multiple cancer clusters with distinct gene expression profiles and spatial distributions. Using an alpha shape approach, we marked cancer-associated fibroblasts and identified upregulated genes in these cells using our MERFISH panel as well as Tangram imputation of genome-wide single-cell expression data. Our analysis revealed the upregulation of genes related to extracellular matrix remodeling and tumorigenesis. These findings illustrate the deep insights into the complex heterogeneity observed in the tumor microenvironment that can be revealed by using the MERSCOPE Platform for oncology research.

SCALE biosciences

The Single Cell Era, Scaled. Finally.

At ScaleBio, our single-cell profiling and analysis solutions based on combinatiorial indexing enable increased throughput for larger, more complex studies.

Achieve an output of 500,000 cells per run at the fraction of the cost.

Millions of combinations never looked so good.

Choose beads and create your own unique bracelet from endless combinations.

Visit us at Suite Bonaire 6, fill out our on-line form for the opportunity to win!

Visit us at Suite Bonaire 6

Informatics & Computational Biology

A collection of software tools to manage illumina sequencing runs

Alberto Riva, Yanping Zhang, Diansy Zincke, Steven J Madore

Interdisciplinary Center for Biotechnology Research
University of Florida, Gainesville, FL USA

Planning sequencing runs on high-throughput Illumina sequencing platforms such as the NovaSeq in a high-volume NGS facility often involves combining multiple pools of samples, usually from different projects, onto the same flowcell. This requires ensuring that there are no barcode conflicts between projects sequenced on the same lane, and generating a single samplesheet by merging the ones for the individual projects. Moreover, samplesheets from different sources need to be manipulated to enforce consistency, or to ensure that barcodes are in the correct orientation. Performing these operations manually is tedious and error-prone, and poses unacceptable risks considering that a single undetected barcode conflict can compromise a whole lane. After the sequencing run is complete, the projects need to be demultiplexed independently, since they may use different barcoding schemes. Finally, it is important to generate reports describing the demultiplexing results in a clear and easily readable format, so that the output of the sequencing run can be immediately assessed.

We have developed a suite of software tools that help all stages of this process. The tools, designed to run in a cluster environment, include: a sample sheet manager that is able to combine multiple sheets into a single one, checking for barcode conflicts, and to perform common manipulations such as reverse-complementing barcodes; a parallel demultiplexing utility that demultiplexes all projects in parallel, performing quality control on the resulting fastq files automatically; and a reporting tool that generates a web page containing the results of demultiplexing (number of reads generated for each lane, for each project, for each sample) in graphical and tabular format.

Together, these utilities make it possible to implement fully automated setup and processing of complex Illumina sequencing runs involving a large number of heterogeneous projects, of the kind typically handled in large NGS facilities.

A telomere-to-telomere reference genome improves read mapping and variant calling in a cross-sectional Swedish cohort

Daniel Schmitz[1], Adam Ameur[1] and Åsa Johansson[1]

1 Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Sweden

Rare Mendelian diseases are often caused by a single genetic variant and together they affect between 3% and 6% of the population. However, even with advanced technologies for whole genome sequencing (WGS) less than half of the affected receive a diagnosis. T2T-CHM13 completes the 8% of the human genome missing from GRCh38, which include functionally relevant regions, such as large parts of chromosome Y. However, it is unclear to what extent T2T-CHM13 would improve the results for population-scale WGS projects and clinical diagnostics. In this study we aim to evaluate the increment of mapping quality and variant call performance with the T2T-CHM13 as a reference compared to GRCh38. We used a cross-sectional Swedish cohort (SweGen) comprising 1000 individuals for which WGS had been performed using Illumina technology. Mapping and variant calling was performed using the established nf-core/sarek pipeline. We observed a higher fraction of reads that could be successfully mapped and properly paired to T2T-CHM13, as compared to GRCh37 and GRCh38, and a lower mismatch rate, giving higher confidence in the mapping results. The number of ambiguous alignments was higher, indicating that alignments appearing unique in older references were not, and coverage of gene regions was more uniform. In comparison to GRCh38, we identified 16.7 million additional variants, including 14.7 million singletons, across the cohort, and observed an increased sensitivity for rare variants. Ensembl Variant Effect Predictor assigned impact ratings of moderate or high for 30% more variants in T2T-CHM13. The average number of variants per individual decreased by ~347,000, driven by a lower number of common variants in T2T-CHM13, possibly explained by variants in GRCh38 that are falsely called in duplicated regions. T2T-CHM13 improved alignment metrics and increased sensitivity for rare and singleton variants and potentially functional variation in SweGen. The improved quality of the mapping and variants calling with the T2T-CHM13 reference genome will enable the discovery of unknown disease-causing variants, including non-coding variants, and insights into variation previously unavailable to precision medicine and genetic epidemiology.

Analysis of spike protein mutations among newly emerging coronavirus subvariants reveals change in mode of entry and establishment of infection in hosts with respect to comparison with previous variants

Asmaa Awan, Roshan Paudel

Morgan State University, SCMNS, Baltimore, MD, USA

Amidst the COVID-19 pandemic, a study on coronaviruses, including their classification and the several strains of the disease-causing pathogen SARS-CoV-2, was conducted. We focused on the strains classified as the variants of concern (VOC) or variants of interest (VOI) or variants under monitoring (VUM) by the World Health Organization (WHO). We have analyzed fifteen different variants and subvariants of SARS-CoV-2, including the newest Omicron subvariant Eris (EG.5.1), with a focus on the spike protein. Our study also looked into the ACE-2 receptor, that allows these viruses to enter human cells via interaction with its surface glycoprotein. A table spanning the full length of the spike protein and other tables conducting pairwise analyses of shared mutations were built.

The most notable mutations observed were D614G, N501Y, E484K/Q/A, P681H and H655Y. The mutations involved with increased transmissibility and replication of the virus by virtue of increased binding capability to the ACE2 receptor are D614G (found in all the variants), N501Y (found in seven out of fifteen variants) and E484K/Q/A (found in 9 out of fifteen variants). On the other hand, the mutations enabling the virus to dodge neutralization by natural and vaccine derived antibodies include N501Y and P681H (found in eight out of fifteen variants). Similarly, the mutation responsible for the change in mode of entry and establishment of the infection by newer variants including Omicron and Eris is H655Y. When mutation in one amino acid is correlated with the probability of mutation in another amino acid in the sequence, it can be assumed that the two amino acids are in close proximity within the protein structure. This study can help understand how different variants are capable of establishing infections in the host differently and need to be targeted by the vaccines accordingly, especially when it comes to the study of the newly emerging variants like Omicron and its subvariants.

Applying transformer-based models to improve splice-junction detection in RNA-seq alignment

Fadel Berakdar, Ankit Sethia, Mehrzad Samadi, Pankaj Vats

Advancement in high-throughput sequencing technologies have revolutionized transcriptomics, enabling researchers to gain valuable insights into gene expression and splicing events. However, RNA-seq data analysis poses significant challenges due to the presence of complex splice-junctions and multi-mapped reads, besides gene fusion events. In this study, we present a novel approach to improve RNA-seq aligners using splice-junction scoring predicted by transformer-based deep learning models.

The primary research question addressed in this study is how to improve RNA-seq aligners' ability to accurately detect splice-junctions and carefully resolve the complex multi-mapped splicing patterns. Our splice-junction scoring approach focuses on leveraging the unique capabilities of transformer models to capture intricate sequence patterns hidden around the donor and acceptor sites.

We fine-tuned the transformer models using GENCODE and RefSeq gene annotation with several window sizes around the donor/acceptor sites. The fine-tuned model is then used as a classifier to predict the splice-junctions found in a BAM file. To benchmark our method, we conducted large-scale RNA-seq alignment benchmark using datasets obtained from Audoux, J et al and Baruzzo et al, which include simulated RNA-seq datasets from human and malaria with increasing levels of complexity (T1 (0.001, 0.0001, 0.005), T2 (0.005, 0.002, 0.01), T3 (0.03, 0.005, 0.02) substitution, indels and error rate respectively. These datasets were then aligned using existing state-of-the-art RNA-seq aligners and each alignment BAM file was processed to predict the reported splice-junction, before applying our transformer-based scoring scheme.

Our preliminary results show improved accuracy while maintaining a similar level of sensitivity in detecting splice-junctions. For instance, benchmarking splice-junction detection in Human T3 dataset aligned by GSNAP version 2023-07-20 shows 98.54% sensitivity and 69.90% accuracy. However, after applying our approach, the sensitivity to 98.87% and the accuracy increased to 91.63%. Similar pattern was observed in the other T1 and T2 Human datasets.

Cell-type deconvolution with long-read, single-molecule methylation

Luke Morina[1], Courtney Hall[1], Jessica Hosea[1], Paul W. Hook[1], Roham Razaghi[1], Winston Timp[1,2]

1 Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, 21218;
2 Department of Molecular Biology and Genetics, Johns Hopkins University, Baltimore, MD, 21218;

Unlike traditional sequencing-by-synthesis technologies, single-molecule platforms (PacBio/ONT) distinguish covalently modified nucleotides directly. With these long (10kb+) single molecules, we can phase epigenetic information across >85% of the genome, enabling allele-specific differential methylation analysis. However, epigenetic characterization of tissue is limited by the complex distribution of cell types. This can lead to either different cell mixtures incorrectly identified as sample-specific differentially methylated regions (DMRs), or actual differential methylation lost as it only occurs within a subset of cells. Profiling cell-type distribution in a tissue sample will allow for improved DMR identification among other downstream epigenetic analysis, especially in known regions of cell-type-specific methylation from bisulfite sequencing. Single-cell sequencing methods attempt to profile this heterogeneity, but often result in sparse data with gaps per cell. Single-molecule long-reads could potentially return cell-type profiles without gaps.

We are developing a toolset using long, single molecules to interrogate the heterogeneous distribution of cellular states within tissue samples by applying it to an in silico mix of well-studied human blood and colon cells. We calculate Euclidean distance between individual reads using either canonical discrete status (unmodified/uncertain/modified) or modification probabilities of individual CpG sites; low-confidence and missing modification calls are accounted for and smoothed within a genomic window size in this process. We are also able to smooth CpGs within a window on a per-read basis. By applying hierarchical clustering of individual DNA molecules using their modification state we tested molecular deconvolution, using the origin of the reads (lymphoblastoid HG002 versus colon epithelial HCT116) as a ground truth. We then applied these same strategies to brain data (from the CARD consortium) and human whole blood, finding that by using known differentially methylated regions (e.g. TREM2 for glia vs. neurons), we identified different cell types. Ultimately, we hope to improve the viability of single-molecule methylation clustering and its applications in epigenetic analysis.

DRAGEN™ Build-Your-Own multigenome (Graph) reference from assemblies: The Human Pangenome Reference Consortium case study

Severine Catreux[1], Massimiliano Rossi[1], Cooper Roddey[1], Ali Chemkhi[1], Mike Ruehle[1], Jennifer Del Giudice[1], Guillaume Rizk[1], James Han[1], Rami Mehio[1],Roham Razaghi[1], Winston Timp[1,2]

1 Illumina Inc, San Diego, CA

Since the release of the first human reference genome, bioinformatic pipelines have relied on linear references for aligning reads in variant calling pipelines. However, linear references cannot represent population genetic diversity, possibly introducing ethnicity biases. The increasingly adopted solution is to use multiple genomes from unrelated and diverse individuals as reference. Therefore, pangenome consortia such as the Human Pangenome Reference Consortium (HPRC) and the Chinese Pangenome Consortium (CPC) are producing high-quality haplotype-resolved genome assemblies of diverse population to establish inclusive pangenome references.

DRAGEN™ supports a multigenome reference since version 3.7.5, that was further extended in version 4.2.4 including more genetic diversity thereby enabling a roughly 50% small variant calling error reduction when compared to the linear reference. The 4.2.4 multigenome reference is built per reference genome using a 32 samples panel provided as phased VCFs, from which a set of optimized haplotypes are derived to improve mapping.

To support the efforts of pangenome consortia in reducing ethnicity bias and democratizing access to genomics, DRAGEN now supports the construction of the multigenome reference directly from haplotype-resolved assemblies. This allows the customization of DRAGEN multigenome references for population-specific studies, or target regions of the genome, e.g. MHC.

We demonstrate the performance of the DRAGEN multigenome construction pipeline by building the multigenome reference using 44 HPRC samples. When compared to dipcall, Minigraph-Cactus, and PGGB as methods to produce phased VCFs from assemblies, the DRAGEN approach resulted in 33% fewer errors than the second-best method, in terms of small variant calling accuracy on the Genome In A Bottle (GIAB) samples aligned to the DRAGEN reference built using the phased VCFs.

(continue to page 70)

(continued from page 69)

Furthermore, the variant calling accuracy using the HPRC DRAGEN reference on the GIAB samples was nearly equivalent (within 4% errors difference) to the accuracy of the optimized 32 samples reference. Finally, we show that we continue to get further accuracy gains by increasing the size of our panel to more than 100 samples. This very promising result suggests that multigenome-based variant calling methods not only improve accuracy, but will have the potential to further reduce ethnicity bias as additional high-quality assemblies are made available.

Enhancing peak calling accuracy by integrating information from multiple biological replicates

Gautam Naishadham, V K Chaithanya Ponnaluri, Sanda Zolj, Heather Raimer Young, Louise Williams, Laura Blum, Matthew A. Campbell, Bradley W. Langhorst

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

Many methods have been developed to produce summary "peaks" for sequencing approaches that produce regions of higher read density relative to diffuse background signal (e.g., ChIP-seq, ATAC-seq, NicE-seq, NEAT-seq, NEED-seq). In a recent review by Eder and Grebien [1], the venerable MACS2 was shown to be a strong contender for general use despite its inability to take advantage of multiple replicates during peak calling. Some analysis methods (e.g., IDR [2], MSPC [3]) use consistency of peaks called between replicates to reject false peaks. This approach requires multiple executions of the peak caller, with each subject to the full noise in the dataset. We reasoned that reducing noise in the original data could provide more information to the peak caller, improving results. We observed a relationship between the inter-replicate coverage variability and the mean signal, suggesting a read filtering approach to provide peak callers like MACS2 with improved signal to noise. The simple method presented here samples reads in replicate bam files by identifying template molecules for exclusion from downstream analysis. Reads are excluded with increased probability as a function of replicate consistency. In comparison with post-calling concordance approaches, this method produced peak calls with better consistency between low and high input libraries, even in challenging FFPE treated samples.

1. Eder, Thomas, and Florian Grebien. "Comprehensive Assessment of Differential ChIP-Seq Tools Guides Optimal Algorithm Selection." Genome Biology 23, no. 1 (May 24, 2022). https://doi.org/10.1186/s13059-022-02686-y.
2. Li, Qiang, James B. Brown, Haiyan Huang, P. J. Bickel, and Peter J. Bickel. "Measuring Reproducibility of High-Throughput Experiments." The Annals of Applied Statistics 5, no. 3 (2011): 1752–79. https://doi.org/10.1214/11-aoas466.
3. Jalili, Vahid, Marzia A. Cremona, and Fernando Palluzzi. "Rescuing Biologically Relevant Consensus Regions across Replicated Samples." BMC Bioinformatics 24, no. 1 (June 7, 2023): 240. https://doi.org/10.1186/s12859-023-05340-x.

Enhancing somatic mosaic variant detection with the DRAGEN™ pipeline

Gavin Parnaby, Yina Wang, Massimiliano Rossi, Arun Visvanath, Lisa Murray, Wei-Ting Chen, Severine Catreux, James Han, Rami Mehio,

1 Illumina Inc, San Diego, CA

Somatic mutations underlying mosaicism occur during embryonic development and throughout an individual's lifetime and have been associated with more than 100 human disorders including cancer and neurological disease. Mosaicism has garnered renewed interest with research efforts aimed at using the latest sequencing technology and analysis tools to deepen our understanding of somatic mosaicism's role in human biology. Progress in mosaicism research is hampered by the lack of bioinformatics methods to detect mosaic variants accurately. Existing bioinformatics approaches used for germline variant calling focus on high allele frequency variants and have low sensitivity for mosaic events. Systematic sequencing artifacts and germline mutations complicate mosaic calling causing false positives. Somatic variant callers designed for oncology applications require background noise files from healthy individuals or have been optimized for paired tumor-normal analysis.

We introduce a mosaic detection algorithm within the DRAGEN germline pipeline that uses an advanced machine learning model to detect and assign quality scores to SNP and indel mosaic variants. It exploits the DRAGEN multi-genome reference to recover calls in low-mappability regions without matched controls. Lower AF calls are recovered compared to other mosaic callers. Mosaic calls are integrated into a standard VCF output alongside germline variants, using tags to ease interpretation.

Our approach was benchmarked using a genotyped cell line admixture dataset and compared to the mosaic callers DeepMosaic and MosaicForecast. Results demonstrate significant accuracy gains across variant allele frequency (VAF). The DRAGEN mosaic caller, benefiting from a larger engineered feature set than other callers, achieved 89.7% recall and 99.1% precision compared with 34% recall and 94% precision for DeepMosaic and 39.4% recall and 94% precision for MosaicForecast. The DRAGEN mosaic caller achieved >90% recall for variants >3.75% VAF in the admixture datasets. The algorithm executes orders of magnitude faster than existing mosaic detection methods. A 40x whole genome can be called in 22.3 mins. This allows mosaic detection to be a default part of analysis pipelines, enabling large scale population studies to uncover new correlations between disease and genomic variation.

Evaluation of splice junction detection methods for single-cell expression data

Edoardo Giacopuzzi, Davide Bolognini, Azahara Fuentes-Trillo, Nicole Soranzo, Cecilia Dominguez Conde.

Humant Technopole, Viale Rita-Levi Montalcini 1, Area MIND, Milan, ITALY.

The detection of splice junctions and isoforms from single-cell gene expression data can offer new insights in the field of single-cell biology, allowing researchers to address unexplored areas of single-cell biology. For example, it sheds light on the regulation of alternative splicing during development, cellular differentiation, and disease cell states. However, splice junction quantification is currently not performed in most scRNA-seq pipelines processing 10x Genomics data, which represents the vast majority of available studies. Despite junctional reads being present in aligned data from 10x Genomics, there is an ongoing debate regarding whether these datasets can be used for reliable splice junction detection and quantification. In this study, we evaluated the performance of different aligners and downstream computational methods for detecting and quantifying splice junctions and isoforms from 10x Genomics gene expression data. Using publicly available data and in-house produced data from both 3' and 5' libraries, including data generated using long-read sequencing, we compared different computational approaches to quantify annotated splice junctions at the single-cell level. We analyzed technical factors influencing splice junction analysis, and we further explored how the resulting splice junction data can be used together with canonical scRNA expression data to analyze differential isoform usage across different cell types and cell states. Our results show that even if difficult at the single-cell level, the comparison of annotated splice junction usage is possible at the cell-type level. Our work addresses current opportunities and challenges associated with splice junction detection from 10x scRNA data and proposes standards for single-cell splice junction analysis. The presented workflow will be available as a dedicated Python package to facilitate future analyses. This creates a great opportunity to re-analyze the massive number of scRNA-seq datasets currently available to the research community.

From development to deployment: Supporting the bioinformatics needs of cutting-edge genomic methodologies and instrumentation

Benjamin J. Kelly[1], Saranga Wijeratne[1], Alejandro Otero Bravo[1], Kelli Roach[1], Jenell Lewis[1], Peter Chang[1], Jocelyn M. Lucyshyn[1], Maria E. Hernandez Gonzalez[1], Anthony R. Miller[1], Elaine R. Mardis[1], Richard K. Wilson[1]

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

In the face of ever-evolving genomic technology, bioinformatics teams supporting these efforts must be agile, innovative, and efficient. To meet those needs, the Computational Genomics Group at the Institute for Genomic Medicine (IGM) is responsible for the bioinformatics development, data analysis, and computational infrastructure necessary to pilot new technologies and validate their use in pediatric research and clinical settings.

Claret Biosciences' SRSLY protocol is a laboratory method ideal for FFPE and other degraded samples, requiring less input and salvaging more DNA molecules for library construction. In addition to utilizing unique molecular identifiers (UMI) for deduplication, novel bioinformatics strategies had to be developed to handle an increase in damage and repair artifacts due to an increased sequencing of molecules that other library preparation methods would have otherwise discarded. In clinically validating this lab-to-bioinformatics workflow, we increased the range of acceptable samples and allowed for more pediatric cancer patients to enroll onto clinical trials.

Liquid biopsy in the pediatric cancer setting has significant advantages in monitoring patients for evidence of recurrence. IGM's translational protocol includes a custom panel per patient, targeting previously identified somatic variants, and a low-pass genome for previously known CNVs from tumor profiling. Adapting our DNA-Seq workflow to accommodate custom loci, taking advantage of high coverage and UMIs to create accurate consensus reads for low-frequency variant identification, and utilizing control-based variants to assess our per-capture sensitivity and precision have provided a personalized and comprehensive, yet efficient, test destined for clinical implementation.

(continue to page 75)

(continued from page 74)

Novel and comprehensive algorithms for characterizing complex isoforms from PacBio IsoSeq data and for processing custom concatenation sequencing methods that allow for more efficient and cost-effective isoform representation have been developed and deployed across our translational research protocols. These methods offer unique transcriptomic insights into disease mechanisms for researchers and clinicians.

In our mission to utilize genomics to help children with cancer and genetic diseases, new technologies, new methodologies, and new instrumentation are critical to finding answers. Rapid laboratory evaluation and adoption require agile bioinformatics support, working together to deliver accurate diagnoses, understand prognoses, and provide insights into treatments and clinical trial options for kids everywhere.

GenCompass: Scalable, cost-efficient and accelerated workflow for unraveling the role of germline genetics in population-level whole-genome sequencing (WGS) studies.

Komal Jain[1], Benjamin T. Jordan[1], Laura E. Egolf[1] and Belynda D. Hicks[1]

1 Cancer Genomics Research Laboratory, Frederick National Laboratory for Cancer Research, Rockville, MD, USA

Technological advancements in next-generation sequencing (NGS) have resulted in increased throughput and reduced cost of sequencing assays, enabling the generation of human WGS datasets for population-level studies. However, analytical pipelines still lag in their ability to process large and complex datasets, often relying on piecemeal solutions generated using a medley of software and smaller pipelines. There is a need for scalable, cost-effective and end-to-end solutions that provide speed and robustness through a careful choice of tools to study the contribution of germline genetics in multiple disease areas. To bridge the gap between data generation and analysis, we have developed GenCompass (Germline Ensemble Calling of Mutations with Parabricks Accelerated Software Suite) - a modular workflow offering comprehensive pre-mapping quality control (QC), accelerated mapping and variant calling using Nvidia Parabricks software, joint genotyping, variant annotation and user-friendly reporting. Additional capabilities of the pipeline include detailed pre- and post-mapping QC metrics, inter- and intra-species contamination, sex, ancestry and sample relatedness checks. In addition to being a complete solution, GenCompass stands out in its ability to perform ensemble variant calling with HaploytpeCaller, DeepVariant and Strelka2 thereby improving precision significantly. We have thoroughly performed benchmarking with individual and jointly called HG001-HG007 samples from Genome in a Bottle and synthetic-diploid samples (Li et al. 2018, Nature Methods) using the latest guidelines. In the joint genotyped benchmarking dataset, GenCompass achieved a median F1 score of 98.4% for SNPs (median precision 99.6%; recall 97.2%) and a median F1 score of 97.7% for indels <50 bases (precision 98.8%; recall 96.5%) at the genome level. By optimizing the runtimes and computational requirements at each step, GenCompass can process 40X WGS samples within 3-4 hours with a computational cost of $7 on GCP using Nvidia Tesla T4 GPU and CPU preemptible instances. Our workflow does not rely on proprietary software and hardware, is system agnostic and uses standardized Docker images, GitHub, Workflow Development Language, and Cromwell for reproducibility. A flexible code design and architecture allows GenCompass to be extended for primary and secondary analysis of a variety of NGS applications.

Genome In A Bottle: New era of assembly-based variant benchmark sets

Nathan D Olson[1], Nathan Dwarshuis[1], Jennifer McDaniel[1], Justin Wagner[1], Justin M. Zook[1], and the Genome In A Bottle Consortium

1 Material Measurement Laboratory, National Institute of Standards and Technology, 100 Bureau Dr, MS8312, Gaithersburg, MD 20899, USA.

Genome in a Bottle was started over 10 years ago with the goal of developing well-characterized human genome reference materials for validating genome sequencing and variant calling methods. GIAB has produced 5 genomic DNA reference materials as well as benchmark sets which are composed of high confidence variant calls as well as regions confidently identified as homozygous reference or as high confidence variants. Currently there are 8 whole genome benchmark sets, small variant benchmarks for 7 individuals including members of an Ashkenazim and Han Chinese trios along with a structural variant benchmark for one of the individuals. These benchmark sets are widely used for evaluating variant calling methods as well as training machine learning and deep learning variant callers. Traditionally these benchmark sets are developed through the integration of multiple variant calls generated using different sequencing technologies and variant calling algorithms. However, this process involves mapping reads to a reference, which is limited in its ability to characterize more complex regions of the genome. With recent advances in sequencing methods and genome assembly algorithms it is now possible to create high quality diploid de novo genome assemblies. We have developed a new Snakemake-based pipeline, Defrabb (development environment for assembly-based benchmarks), that generates high quality whole genome small and structural variant benchmark sets from assemblies. These benchmarks include regions of the genome that were excluded when using read-mapping-based benchmark generation methods, such as the highly polymorphic MHC, gene conversions, and more complex SVs. Here we present our new pipeline along with our first X and Y benchmark set and a draft whole genome SV benchmark set. The XY benchmark covers 94% of X and 63% of Y, with 87,452 SNP and 24,273 indels. The draft HG002 SV benchmark for GRCh38 covers ~2,788,000,00 bases with nearly 30,000 variants. These new assembly-based variant call benchmark sets will allow for the validation of challenging, previously unbenchmarkable variant calls and can be used to train deep learning models to more accurately call variants in these challenging regions.

Integrating CapTrap-Seq reads into the GENCODE geneset

Toby Hunt[1], Jose Gonzalez[1], Silvia Carbonell-Sala[2], Gazaldeep Kaur[2], Jonathan Mudge[1], Roderic Guigó[2], Adam Frankish[1], on behalf of the GENCODE consortium.

1 European Bioinformatics Institute, European Molecular Biology Laboratory, Cambridge, United Kingdom.
2 Centre for Genomic Regulation, the Barcelona Institute of Science and Technology, Barce lona, Spain.

The GENCODE consortium produces detailed reference annotation of all human and mouse protein-coding genes, pseudogenes, long non-coding RNAs and small RNAs. Accurate gene annotation is of fundamental importance for genome biology and clinical genomics; annotation that is incorrect or incomplete impacts downstream analysis and introduces potentially significant errors.

As part of the consortium's efforts, the CapTrap-seq protocol has been specifically designed to achieve reliable end-to-end production of long-read RNA sequences using both PacBio and ONT platforms. Combining CapTrap and oligo(dT) priming to detect 5'-capped, full-length transcripts with the Capture Long Seq (CLS) method we have been able to identify more than 100,000 novel human lncRNA transcripts. These highly accurate transcripts, whose introns are verified by short-read RNAseq data, can then be integrated into the GENCODE geneset via our automated TAGENE pipeline with minimal manual oversight.

Thus we are continuing to refine and improve the GENCODE geneset in multiple ways. 1. By incorporating this new long-read transcriptomic data to extend existing models and add new ones. 2. Identifying novel ORFs, including potential translations within lncRNAs and the UTRs of protein-coding genes, found within this new annotation and supported by recently published Ribo-seq studies. 3. Via the production of high-value sets of functionally defined transcripts such as the Matched Annotation from NCBI and EMBL-EBI (MANE) produced in collaboration with NCBI RefSeq, and the newly available GENCODE Primary subset of our GENCODE Comprehensive gene annotation.

The GENCODE genesets are the default Human and Mouse annotation used in the Ensembl and UCSC genome browsers and can be downloaded from www.gencode-genes.org.

Leveraging the power of fine-scale population structure to identify clinical and social determinants of disparities in sepsis risk

Abhijith Biji

Background: Sepsis, a life-threatening dysregulation of immune response to infections, disproportionately affects African/Americans (AA) and Hispanic/Latinos (HL) relative to European Americans (EA). Epidemiological studies have linked this to socioeconomic status which can be a proxy for variables such as lifestyle, access to care, income, or clinical risk factors. Moreover, commonly used race/ethnicity (R/E) population labels may not adequately represent variation between sub-populations within R/E groups. Addressing this is essential for identifying risk factors involved in sepsis disparity.

Methods: We use genetic data, sociodemographic data and electronic health records from the BioMe biobank (68% non-white). Our cohort includes 1360 sepsis cases (395 AA, 547 HL, 282 EA) and 47038 controls (9242 AA, 15753 HL, 15019 EA). Then, using a novel method based on recent shared genetic ancestry (Belbin GM et al, Cell, 2021), we identified fine-scale population groups (eg, Dominican (DOM) or Puerto Rican (PR) instead of broad self-reported HL). Finally, we assessed sepsis clinical risk factors in each fine-scale population group.

Results: Sepsis had a higher odds ratio (OR) in AA (OR 2.61[95%CI 2.23-3.05], p-value (p)=4.18e-33) and HL (2.06[1.78-2.38], p=6.1e-22) relative to EA. Moreover, younger (<50 yr) AA and HL individuals faced even higher sepsis risks than same age group EA (2.95[1.92-4.53]) and 3.04[2.04-4.54] respectively). Comparing clinical risk factors, we saw younger vs older AA had higher sepsis OR for 214 conditions including infection from methicillin-resistant Staphylococcus aureus and respiratory failure, and younger vs older HL had higher sepsis OR for 286 conditions including renal failure, infection with drug-resistant microorganisms and respiratory failure. Among fine-scale population groups, PR and DOM exhibited distinct sepsis OR (2.13 and 1.63 respectively), a nuance missed when using self-reported ancestry alone. Notably, we also observed that clinical risk factors of sepsis differ between them; diseases related to kidney function had a higher odds ratio in DOM while urinary tract infection was a bigger risk factor in the PR.

Conclusion: Leveraging EHR-linked genetic data from large-scale biobanks allows us to identify population-specific risk factors of sepsis that cannot be identified using traditional epidemiology and can pave the way toward targeted preventive and public health policies.

OMKar: Optical Map based automated karyotyping of genomes to identify constitutional abnormalities

Siavash Raeisi Dehkordi[1], Joey Estabrook[2], Zhaoyang Jia[1], Jen Hauenstein[2], Neil Miller[2], Alex Hastie[2], Andy Wing Chun Pang[2], Vineet Bafna[1]

1 Department of Computer Science and Engineering, University of California San Diego, San Diego, CA, USA;
2 Bionano Genomics, San Diego, CA92121, USA;

The whole genome karyotype refers to the sequence of large chromosomal structural aberrations that make up the genotype of an individual, covering variants such as aneuploidies, balanced and unbalanced translocations, and others. Many such anomalies have been implicated in constitutional genetic disorders. Karyotyping is crucial for diagnosis, treatment decisions, and genetic counseling, providing foundational insight into chromosomal health and potential genetic risks. Optical genome map (OGM) based investigation has the potential to generate virtual karyotypes, reducing the reliance on visually microscopic examination.

We developed OMKar, which leverages OGM data to construct a virtual karyotype. OMKar accepts Structural (SV) and Copy Number (CNV) Variants as inputs. Briefly, it first eliminates low-confidence calls, and generates a breakpoint graph where each genomic segment is symbolized by dual vertices, and edges join adjacent segments or structural breakpoints. Applying Integer Linear Programming, OMKar smooths edge multiplicities, ensuring CN balance condition which guarantees edge sums remain within segment multiplicity bounds. Subsequently, it identifies and scores Eulerian paths within the breakpoint graph, where each path represents a probable karyotypic structure. We tested OMKar on 10 simulations. In each case, we simulated the karyotype of a known constitutional disorder and added 10 through 15 additional SVs. OMKar identified the constitutional aberration with an accuracy of 87%, while also identifying SVs with high precision (87%) and recall (94%).

We applied OMKar on 52 prenatal samples, including 3 control samples and others with a genetic syndrome. OMKar was able to detect a cytogenetically validated karyotype in 45 of the 49 cases, including 29 of 29 balanced translocations, 4 of 4 complex events with multiple abnormalities, and 11 of 16 aneuploidies. The missed aneuploidies often occurred for mosaic aneuploidies. The detected aneuploidies included many known syndromes such as Cri-du-chat, Wolf-Hirschbolm, Prader-Willi deletions, Down's and Turner's syndromes. Together these results point to the robustness of OGM karyotyping using OMKar.

Resolving and cataloging variation in long and polymorphic regions of the human genome

Egor Dolzhenko[1], Graham S Erwin[2], Katherine Wang[2], Zev Kronenberg[1], William J. Rowell[1], Anna C Ferrari[3], Garrison Pease[4], Daniel Schwartz[4], Benjamin Gartrell[3], Ahmed Aboumohamed[5], Alex Sankin[5], Pedro Maria[5], Kara Watts[5], John Greally[6], Robert Dubin[6], Patrick Wilkinson[7], Jaymala Patel[7], John Loffredo[7], Denis Smirnov[7], Manuel A. Sepulveda[7], Charles G. Drake[7], Alex Robertson[1], Michael P Snyder[2], Michael A Eberle[1]

1 Pacific Biosciences of California, Menlo Park, California, USA
2 Stanford University, Department of Genetics, Stanford, USA
3 Prostate Cancer Annotated Biobank for Minorities (PCABM) Montefiore-Einstein Cancer Center, Department of Oncology, Bronx, New York, USA
4 Prostate Cancer Annotated Biobank for Minorities (PCABM) Montefiore-Einstein Cancer Center, Department of Pathology, Bronx, New York, USA
5 Prostate Cancer Annotated Biobank for Minorities (PCABM) Montefiore-Einstein Cancer Center, Department of Urology, Bronx, New York, USA
6 Einstein Epigenomics Center, Bronx, New York, USA
7 Janssen Research and Development LLC, Immuno-Oncology Discovery, Spring House, Pennsylvania, USA

The human genome contains thousands of repeat-rich polymorphic regions whose structure has not been systematically described. These regions produce large collections of variant calls sometimes called variation clusters. Variation clusters are typically excluded from tertiary analysis because it is difficult to interpret and catalog them. One example is the 3.5 Kbp region in an intron of the KCNMB2 gene which contains over 30 constituent simple repeats that jointly create many insertions, deletions, and mismatches in alignments of reads over this region. The corresponding variant calls are often incorrectly prioritized as potentially pathogenic, requiring significant resources to curate and rule out. To address these issues, we propose a novel computational framework to systematically detect, annotate, and catalog variation clusters. A distinguishing characteristic of our approach relative to traditional variant calling methods, is the ability to annotate and resolve entire regions of high sequence polymorphism as single units instead of fragmenting them into variation clusters. These regions can be subsequently genotyped using the recently developed tandem repeat genotyping tool (TRGT).

(continue to page 82)

(continued from page 81)

We show that our method can accurately detect reference coordinates and resolve structures of KCNMB2, MUC1, CEL, INS and 20 other medically relevant variable number tandem repeats. Using real and simulated data we also show that our method can locate and call pathogenic expansions of 50 disease-causing repeats and nine polyalanine repeats composed of highly variable motif sequences. To demonstrate the usefulness of our method for cancer genome studies, we applied it to normal, polyp, and adenocarcinoma PacBio HiFi samples originating from the same individual and identified a tandem repeat that progressively expands in length from normal to polyp to adenocarcinoma samples in the 5′ UTR of LIMD1, a reported tumor suppressor gene. To further highlight our ability to resolve variation we characterized differences in repeat composition and methylation between three prostate tumors and their normal counterparts and also a panel of 100 unrelated genomes. To make all these analyses accessible to other genome researchers, we are releasing a learning resource with tutorials describing how to catalog variation in the polymorphic regions of the human genome in publicly available PacBio HiFi samples.

Single-molecule-level analysis discovers DNA motifs that potentially delay DNA polymerase.

Kei Shimazawa and Shinichi Morishita

1 Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, University of Tokyo.

PacBio's single-molecule real-time (SMRT) sequencing outputs the data of the interval time between reading each base, which is called inter-pulse duration (IPD). IPD often increases when DNA modification such as N6-methyladenine is observed; however, there seem to be different causes of the high IPD values. Previous SMRT sequencing, called continuous long reads (CLR), typically scans each DNA molecule and measures IPD once. Due to the nature of single-molecule observations, IPD values for individual locations can vary between different CLR reads, so the average value is instead used as the reliable IPD value. Current HiFi sequencing (or circular consensus sequencing, CCS), on the other hand, can measure IPDs of the same molecule multiple times, resulting in more reliable IPD values. Thus, HiFi reads are expected to help understand the relationship, at the single molecule level, between the context and the high IPD values without known modification. We investigated the HiFi long-read sequencing data of C. elegans and attempted to find bases with consistently and significantly high IPD for each CCS read using binominal testing. The result seems congruent with the previous study with CLR since the number of detected bases within the motifs presumably causing the formation of non-B DNA structures is 3-4 times greater than that of other random motifs. Moreover, some new motifs, which seemed undiscovered in the previous CLR studies, are reported as having a high detection rate to the coverage, such as the AT repeat-like region or the one with GDMTGARCA. This research contributes to comprehending the mechanism of long-read sequencing, especially the correlation between IPD fluctuation and sequence contexts.

SpatialGPT: Extending pre-trained Transformer Foundation Model in spatial transcriptomics and proteomics

Jiwoon Park[1], Roberto De Gregorio[1], Brian Robinson[1], Erika Hissong[1], Sanjay Patel[1], Fabio Socciarelli[1], Patrick Danaher[2], Evelyn Metzger[2], Liuliu Pang[2], Nathan Schurman[2], Jason Reeves[2], Yan Liang[2], Joachim Schmid[2], Raymond Tecotzky[2], Sajad Darabi[3], Harry Clifford[3], Steven Kothen-Hill[3], George Vacek[3], Joseph Beechem[2], Massimo Loda[1], Olivier Elemento[1], Alicia Alonso[1], Shauna Lee Houlihan[1], Robert E. Schwartz[1], Christopher E Mason[1]

1 Weill Cornell Medicine, New York, NY, USA
2 NanoString® Technologies, Seattle, WA, USA
3 NVIDIA Corporation, Santa Clara, CA, USA
*Corresponding author: Christopher E. Mason

Machine learning has undergone a paradigm shift with the rise of models trained on extensive, broad data using self-supervision at scale, which can be adapted to a wide range of tasks. For example, single-cell foundational models can be used to extract a rich representation of any single-cell profile, which can power cell type prediction and batch correction. They can also serve as a base model that can be finetuned with limited training data to perform other prediction tasks, such as predicting the impact of perturbations on cells.

While single-cell methods have generated diverse datasets from nearly every human tissue, they are limited by isolation biases, sequencing methods, batch effects, and computational pipelines. These datasets also lack spatial information such as tissue architecture, cellular neighborhoods, and intercellular signaling, which contextualize how cells engage in situ. To address these limitations, we have extended a single-cell transformer-based foundation model into spatial transcriptomics and proteomics data, specifically from the Spatial Atlas for Human Anatomy (SAHA), a collection of diverse tissues profiled on the CosMx Spatial Molecular Imager (SMI) with a universal 1000-plex RNA panel and 64-plex immunophenotyping proteomics panel, as well as datasets collected with a new CosMx 6000-plex RNA panel.

(continue to page 85)

(continued from page 86)

Training foundation models on spatial data presents unique challenges due to the smaller number of features than scRNA-seq and off-target, background counts in cell profiles. However, it also offers new opportunities, such as validating cell types and investigating their behavior in response to nearby surroundings and perturbations. We benchmark standard omics tasks (i.e., cell typing, batch integration, cell-cell interaction, ligand-receptor engagement), evaluate success compared to individual regression models and pathological reviews, and assess the impact of feature size (plex) in spatial data on model performance. We also extend the training and inference infrastructure leveraging NVIDIA's BioNeMo platform to deliver these findings at scale.

Incorporating spatial information into GPT-based foundation models can significantly improve their performance on a variety of tasks in single-cell biology and allow researchers to maximize the use of the generated datasets, which will enable new possibilities for understanding cell behavior and developing new diagnostic and therapeutic strategies.

StratoMod: Predicting sequencing and variant calling errors with interpretable machine learning

Nathan Dwarshuis[1], Peter Tonner[1], Nathan D. Olson[1], Fritz J Sedlazeck[2,3], Justin Wagner[1], Justin M. Zook[1]

1 Material Measurement Laboratory, National Institute of Standards and Technology, 100 Bureau Dr, MS8312, Gaithersburg, MD 20899
2 Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA
3 Department of Computer Science, Rice University, 6100 Main Street, Houston, TX, 77005, USA

The Genome in a Bottle (GIAB) consortium generates variant benchmarks for a set of human genomes to enable evaluation and comparison of sequencing technologies and variant detection methods. While these technologies have advanced significantly, correctly calling variants in complex or repetitive regions remains a challenge. We generally understand that sequencing biases and short read lengths can lead to incorrectly-called variants; however, we lack a data-driven model that uses GIAB benchmarking metrics to link variant caller performance to quantifiable features of the context surrounding a given variant.

We made such a model using explainable boosting machines (EBMs) which fit data using a linear combination of arbitrary univariate and bivariate functions (a generalized additive model with interaction terms). Despite being flexible, the simplicity of EBMs allows a human to easily understand how each feature contributes to the outcome. We exploit this transparency by fitting EBMs to variant call errors as a function of genomic context, which enables understanding how given feature will impact performance.

We demonstrated several uses cases for this method. First, we investigated false positive error rates of Illumina sequencing using PCR-free or PCR-plus libraries; error rates were unsurprisingly higher for PCR-plus, and we identified the minimum length of homopolymer (for SNVs this was 8bp for AT and 4bp for GC) beyond which errors became more prevalent. Second, we compared Illumina and Hifi in their ability to correctly identify clinically relevant variants. The transparent nature of our model allowed us to identify that while many errors occurred in hard-to-map regions (particularly for Illumina), other missed variants were located in homopolymers or tandem repeats. Third, we utilized an assembly-based benchmark based on an HG002 T2T assembly to quantify homopolymer performance between Element and Illumina. Our model in combination with this new benchmark enabled precise quantification for AT and GC homopolymers up to 50bp.

(continue to page 87)

(continued from page 88)

Ultimately, this provides a data-driven mechanism for understanding sources of error within different variant caller methods and sequencing technologies within difficult genomic regions. We also plan to use this model to improve genome stratifications for creating and using GIAB benchmarks. The code is available at https://github.com/ndwarshuis/stratomod.

The NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-space (AnVIL)

Natalie Kucher[1], Michael C. Schatz[1,2], Anthony Philippakis[3], AnVIL Project Team[4]

1 Department of Biology, Johns Hopkins University, Baltimore, MD.
2 Department of Computer Science, Johns Hopkins University, Baltimore, MD.
3 Broad Institute of MIT and Harvard, Cambridge, MA.
4 City University of New York, Harvard University, Carnegie Institution for Science, Oregon Health & Sciences University, Penn State, Roswell Park Cancer Institute, University of California Santa Cruz, University of Chicago, Vanderbilt University, Washington University, American Heart Association, Fred Hutchinson Cancer Research Center

The traditional model of genomic data sharing – using 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 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, 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 as needed. AnVIL currently provides harmonized access to >600,000 genomes, with many more on the horizon. We also provide access to thousands of software tools, plus several options for interactive and batch analysis.

First, we discuss how AnVIL supports the Telomere-to-Telomere (T2T) consortium through a large-scale reanalysis of the 1000 Genomes cohort orchestrated through the Workflow Description Language (WDL) on Terra. This enabled us to identify over one million variants in the newly resolved regions of the human genome, including within the recently completed chromosome Y sequence. Next, we present new results detecting single- and multi-tissue SV-eQTLs by genotyping SVs discovered with long-reads within the GTEx short-read sequencing data. This work was securely executed through WDLs, Jupyter notebooks, and R/Bioconductor, and lead to the discovery of 5,580 SV-eQTLs where the SV has the highest CAVIAR score (a metric of causality) over other nearby SNVs. Finally, we discuss how to analyze pangenomes and haplotype diversity using Galaxy within Terra. This is critically important to diversity and disease studies, especially to capture and analyze variation not found in any single reference genome.

(continue to page 89)

(continued from page 88)

Long-term, AnVIL will offer a unified platform for ingestion and organization for a multitude of 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.

BioSkryb GENOMICS

Join us in **Antigua 1** to learn about
the Next Generation of Single-Cell Technology

Resolve MORE

Get the most comprehensive view
of the genome, transcriptome, and
targeted proteins

Medical Omics

A blood collection workflow for interrogation of translational samples with single-cell technologies

Dagmar Walter[1], Jawad Abousoud[1], Tingsheng Yu Drennon[1], Kelly Martin[1], Connor Kunihiro[1], Yina Li[1], Jill Herschleb[1]

1 10X Genomics, Pleasanton, CA

Harnessing cells that circulate in the bloodstream is a critical step in understanding many complex diseases. Readouts obtained from blood help inform clinical treatments, investigate disease mechanisms, evaluate drug response, and monitor outcomes. However, the impact of existing blood collection workflows on data output from new and high resolution technologies, like single cell RNA sequencing (scRNA-seq) remains a challenge.

Recent technological advances in scRNA-seq assays have seen a transition from reverse transcription primed workflows that require intact RNA, to probe based methods that are compatible with degraded RNA and thus, fixation. Sample fixation has been shown to stabilize samples during transport and is compatible with downstream bulk RNA-seq, however, this approach does not preserve intact cells so is not compatible with single cell technologies. Thus there is a need to preserve blood while retaining single cell information. In this study, we evaluated various PBMC isolation and blood preservation methods which preserve single cell information while easing sample logistical constraints.

Our initial experiments compared various PBMC isolation methods. scRNA-seq analysis of these samples demonstrated that all isolation methods captured single cell gene expression with some variability in cell types along with user-variability. Changes in cell type proportions were observed in as little as 4 hours post blood draw, emphasizing the need for an effective blood preservation method to retain single cells during transport and blood cell isolation.

Next, we compared different methods of whole blood preservation including cryopreservation and fixation. Healthy and diseased blood samples were collected, preserved, stored followed by isolation of blood samples at different time points. scRNA-seq analysis of the fixed samples compared to fresh samples demonstrated that blood can be preserved, transported, and stored (weeks to months) before blood cell isolation, yielding comparable data quality and cell type proportions. Blood cells isolated from these samples can then be stored long term for scRNA-seq analysis. This blood collection and preparation workflow also eases blood transportation logistical constraints, allowing for batch shipping of samples from distributed collection sites. Coupled with automatable cell isolation methods and multiplexing single cell sequencing solutions, these workflows can be adapted for large scale translational research studies.

A highly faceted scientific management model catalyzes large-scale team science to help build the future of genomic medicine.

Lucinda Antonacci-Fulton[1], Tina Lindsay[1], Sarah Cody[1], Milinn Kremitzki[1], Heather Lawson[1], and Ting Wang[1]

Washington University School of Medicine

Washington University School of Medicine has a long history in the field of genomics, contributing to the generation of the original human genome project. It is home to organizational centers that drive large-scale, team-based genomic research.

- Human Pangenome Reference Consortium (HPRC) – Building human pangenome resources will lead the scientific and clinical communities from a linear human reference to a reference comprised of many genomes that better represent human variation, leading to improved outcomes in the clinic.
- Impact of Genomic Variation on Function (IGVF) – Study of genomic variation on function leading to the creation of a catalog of results and resources utilized to understand phenotypes
- Somatic Mosaicism Across Human Tissue (SMaHT) is a systemic study of genetic differences within the body to create a catalog of variations to know how they im pact biology and disease.

HPRC will allow for the development of a multi-allelic reference comprised of many genomes that better define diversity. This new reference landscape and information about somatic variation will allow for a broader investigation of genomic variation on function. The resulting information will collect functionalized germline and somatic variations.

Funded components comprise the consortia: an organizational unit, data producers, analysis core, and groups funded for tool and technology development. Resources are regularly cataloged and made available to the scientific community. Outreach activities include publications, presentations, public-facing websites, and social media. The organizational framework for each consortium consists of a steering committee and working groups focused on specialized work streams.

(continue to page 94)

(continued from page 95)

Wiki software allows for the collaborative organization and management of information and resources generated by the consortium. Google and Box store and retain files. Communication tools enhance daily interactions and involve SLACK and email. Many consortia also have mid-year virtual and annual, in-person meetings that allow the consortia researchers to come together to take stock of their current work and make plans for the future.

Centralizing activities within a multi-consortium support center yields the quick start of consortia, an operational economy of scale, quick adoption of new organizational methods and policy changes, systematic in-reach and outreach strategies, and the straightforward cross-pollination of systems, and ideas between the consortia.

Clinical validation of short tandem repeat Expansion variant calling using short-read WGS data

Andrea Oza[1], Diana M. Toledo[1], Marina DiStefano[1], Areesha Salman[1], Katherine Lafferty[1], Mark Fleharty[1], Edyta Malolepsza[1], Ning Li[1], Alyssa Friedman[1], Samuel DeLuca[1], Teni Dowdell[1], Justin Abreu[1], Fontina Kelley[1], Grace VanNoy[1], Ben Weisburd[1], Heidi Rehm[2,1], Niall Lennon[1]

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

Testing for short tandem repeat (STR) expansions previously required specific assays such as triplet-repeat primed PCR or Southern blot. Recently developed computational tools, such as ExpansionHunter, can now identify STR expansions using PCR-free short-read WGS data to genotype STR loci. The Rare Genomes Project (RGP) at the Broad Institute has tested and implemented ExpansionHunter to identify variants responsible for rare and undiagnosed conditions. To date, 10 individuals with features of ataxia, myopathy, or muscular dystrophy and onset ranging from birth to 81 years have been diagnosed with an STR expansion in 6 different genes (NOTCH2NLC, ATXN1, ATXN2, RFC1, CNBP, PABPN1).

Success of ExpansionHunter in RGP has led the Broad Clinical Laboratory to validate Illumina DRAGEN ExpansionHunter for clinical WGS testing. The validation cohort includes 22 samples with known STR expansions in six genes (FMR1, ATXN1, FXN, HTT, C9ORF72, and DMPK) that vary by repeat size, motif, inheritance pattern, and patient sex. Performance will be analyzed for each sample by comparing the genotyping call in the repeat VCF to the truth data. Repeats smaller than the sequencing read length (~150bp) are expected to be accurately sized (+/-1 repeat); however, the size of STRs that are ~150bp or longer will likely be underestimated (PMID: 35182509). This will be accounted for by allowing STRs >150bp to pass validation if the predicted repeat size is within +/-1 repeat or if the size of the repeat correctly correlates to the gene's normal, premutation, or pathogenic repeat range.

(continue to page 98)

(continued from page 97)

After analytical sensitivity, specificity, and precision criteria are met, additional steps will be taken to incorporate DRAGEN ExpansionHunter calls into a written interpretive report. These include gene curation, annotation of the VCF file to flag STR expansions outside of the normal repeat range, development of a manual review procedure using REViewer, a procedure for orthogonal confirmation, and the procedure for writing a clinical interpretive report.

Validation of STR calling will expand the variant types reported in clinical WGS, increasing sensitivity for neurological disorders that may be caused by STR expansions, and reducing the need for separate test orders for individuals at risk for these disorders.

Genome sequencing as a generic diagnostic strategy for rare disease

Marcel Nelen[1], Gaby Schobers[1,2], Ronny Derks[1], Amber den Ouden[1], Jeroen van Reeuwijk[1,2], Ermanno Bosgoed[1], Dorien Lugtenberg[1], Su Ming Sun[3], Jordi Corominas Galbany[1,2], Marjan Weiss[1], Rien Blok[3], Richelle Olde Keizer[1,2] , Debby Hellebrekers[3], Nicole de Leeuw[1], Alexander Stegmann[3], Erik-Jan Kamsteeg[1], Aimee D.C. Paulussen[3], Marjolijn Ligtenberg[1,2], Xiangqun Zheng Bradley[4], John Peden[4], Alejandra Gutierrez[4], Adam Pullen[4], Tom Payne[4], Christian Gilissen[1,2], Arthur van den Wijngaard[3], Han G. Brunner[1,2], Helger G. Yntema[1,2]*, Lisenka E.L.M. Vissers[1,2]*

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

Background: To diagnose the full spectrum of hereditary and congenital diseases, genetic laboratories use many different workflows, ranging from karyotyping to exome sequencing. A single generic high-throughput workflow would greatly increase efficiency. We assessed whether genome sequencing (GS) can replace these existing workflows aimed at germline genetic diagnosis for rare disease.

Methods: We performed GS (NovaSeq6000™; 37x mean coverage) on 1,000 cases with 1,271 known clinically relevant variants, identified across different workflows, representative for our tertiary diagnostic center. Variants were categorized in small variants (SNVs/indels <50 bp), large variants (CNVs and STRs) and other variants (SVs and aneuploidies). VCFs were queried per variant, from which workflow-specific true positive rates (TPRs) for detection were determined. A GS-first scenario was modeled for 24,570 individuals referred to our laboratory in 2022, using diagnostic efficacy and predicted false negative as primary outcome measures.

Results: Overall, 95% (1,206/1,271) of variants were detected. Detection rates differed per variant type: small variants in 96% (826/860), large variants in 93% (341/366), and other variants in 87% (39/45). TPRs varied between workflows (79-100%), with 7/10 being replaceable by GS. Models for our laboratory indicates that a GS-first strategy would be feasible for 84.9% of clinical referrals, translating to 71% of all individuals receiving GS as their primary test. A possible false negative rate of 0.3% could be expected.

Conclusion: GS can capture clinically relevant germline variants in a 'GS-first strategy' for the majority of clinical indications in a genetics diagnostic lab.

Genomic characterization of BSL-3 and BSL-4 viruses and infectious disease pathogenesis

Monika Mehta, Lirong Peng, Shawn Hirsch, David Drawbaugh, Sanae Lembirik, Saurabh Dixit, Erin Kollins, Russ M. Byrum, Yu Cong and Michael R. Holbrook

1 Integrated Research Facility at Fort Detrick, Division of Clinical Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Frederick, MD, USA

In the post-pandemic world, the imperative of understanding and managing highly contagious pathogens has gained heightened significance. The Integrated Research Facility of NIAID at Fort Detrick, MD (IRF-Frederick) employs a multidisciplinary approach to understand, treat, prevent, and eradicate high-consequence diseases caused by novel, emerging, and highly virulent viruses. The genomics team at IRF-Fredrick utilizes next-generation sequencing techniques to conduct comprehensive genomic analyses of viral pathogens including those that require biosafety level (BSL)-3 and BSL-4 containment, such as Lassa virus, Ebola virus, Nipah virus, and SARS-CoV-2 etc.

Despite substantial technological advances in the development of novel genomics applications in recent years, applying these innovative approaches to study these pathogens has been difficult due to the unique challenges presented by BSL-4 containment. Our team has been working towards testing and optimizing methods for safely generating and sequencing virus and host samples for genomic and transcriptomic characterization using short-read, long-read, and single-cell sequencing. This presentation will delve into some of the challenges encountered while testing and optimizing these protocols for high-containment pathogens and outline the solutions we have employed to overcome those challenges using a set of SARS-CoV-2 studies as an example.

To study the natural history of the various SARS-CoV-2 variants, several collaborative studies were undertaken at IRF-Frederick where the model hosts, Syrian Golden hamsters were challenged intranasally with different SARS-CoV-2 variants and studied for up to 10 days using virological, histopathological, and immunological assays. In order to analyze the host transcriptomic changes during the progression of infection and virus clearance, we used lung samples collected 3, 7 or 10 days after infection and analyzed the host transcriptome using bulk and single-cell RNA sequencing. In addition to temporal analysis of disease progression, this also enabled us to study the similarities and differences between the host response to infections caused by 10 major SARS-CoV-2 variants. Additionally, we are analyzing the changes to the virus genome during the infection cycle. Some of the interesting findings of this study will be presented, along with some preliminary data from ongoing Nipah virus studies.

Interrogating the transcriptional landscape of interstitial lung disease with FFPE single-cell and spatial RNA profiling.

Miles Tran[1], Ce Gao[1], Mukta Palshikar[1], Jessica Liu[1], Daimon Simmons[1], Mari Kamiya[1], Gregory McDermott[1], Deepak Rao[1], Jeffrey Sparks[1], Edy Kim[1], Ilya Korsunsky[1], Kevin Wei[1]

Brigham and Women's Hospital, Department of Rheumatology, Boston, MA, USA

Single-cell RNA sequencing (scRNA-seq) has become the de facto approach for deconstructing complex biological systems and diseases. One recent variation developed by 10X Genomics, single-cell FLEX Profiling (FLEX), allows for highly multiplexed transcriptomic profiling of formalin-fixed paraffin-embedded (FFPE) tissue at single-cell resolution. As FFPE tissue is used in standard clinical pathology laboratory settings, FFPE-compatible chemistries unlock transcriptomics for a vast bank of clinically annotated patient samples. In contrast to traditional scRNA-seq, FLEX enables capture of fragile cell types, retrospective cohort design, and generation of atlas-scale single-cell data sets in a single assay. Additionally, spatial transcriptomics assays have also been modified to use FFPE-compatible chemistries. 10X Genomics's Visium FFPE Assay allows for spatially resolved transcriptomic profiling of FFPE tissue at a super-cellular resolution. Integrating the single-cell resolution of FLEX and the spatial context of Visium offers a powerful new lens to illuminate the histological, cellular, and molecular landscape of disease.

Interstitial lung disease (ILD) encompasses a heterogenous group of disease subtypes broadly characterized by alveolar inflammation and fibrosis of the lung. While the ILD subtypes have a diverse range of clinical outcomes, they often exhibit similar clinical manifestations during early disease. Accordingly, there remains an ongoing search for axes of variation that facilitate robust and timely diagnostic stratification of ILD. One potential avenue for this stratification comes from looking at differential abundance of cell types in diseased tissue with the help of single-cell transcriptomics.

Here, we leverage multiplexed FLEX profiling of over 400k single cells, matched spatial transcriptomic data consisting of over 175k Visium spots, and clinical annotations to deconstruct ILD at the single-cell level with spatial contextualization. Spanning 31 samples, 10 ILD subtypes and dozens of fields of clinical metadata, we demonstrate that these two complimentary assays facilitate the stratification of early ILD into single-cell metaphenotypes that lie along an inflammatory-fibrotic axis. We show that these metaphenotypes, initially characterized at the single-cell level by examining differential abundance of cell types across samples, correspond to relevant histopathological features such as plasma cell proliferation, increased occurrence of lymphoid aggregates, atypical alveolar morphology, and large swathes of myofibroblast rich regions in the tissue.

Leveraging optical genome mapping and RNA-seq to identify new candidates in an x-linked endothelial corneal dystrophy (XECD) family.

Joshua D. Smith[1], Karynne Patterson[1], Jessica X. Chong[2], Colby T. Marvin[2], Kati J. Buckingham[2], Jeff Weiss[1], Danny Miller[2], Miranda Galey[2], Jesse Bengtsson[3], Claudia M. B. Carvalho[3], University of Washington Center for Rare Disease Research, Evan E. Eichler[1,4], Michael J. Bamshad[1,2], Andreas R. Janecke[5]

1 Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA
2 Department of Pediatrics, University of Washington, Seattle, WA 98195, USA
3 Pacific Northwest Research Institute, Seattle, WA 98122, USA
4 Howard Hughes Medical Institute
5 Institut für Medizinische Biologie und Humangenetik, Medizinische Universität Innsbruck, Innsbruck, Austria

X-linked endothelial corneal dystrophy (XECD) is an exceptionally rare form of posterior corneal dystrophy characterized by diffuse corneal haze and congenital ground glass corneal clouding leading to vision impairment. The phenotypic features of XECD vary widely based on gender, with males experiencing more severe impacts than females. Both males and females with XECD display changes in the corneal endothelium that resemble "moon craters." However, advanced cases in males have shown associations with congenital ground glass corneal opacification, nystagmus, subepithelial band keratopathy, and worsened visual acuity. As the least common condition among corneal dystrophies, XECD was identified in a single 7-generation multiplex pedigree from Austria. Utilizing 25 polymorphic markers spread across the entire X chromosome, linkage analysis identified a 14.79 MB region on the long arm of Xp25 associated with XECD and this linkage region was subsequently confirmed by genome-wide chip genotyping and linkage analysis. Within the linkage region, there are a total of 181 genes, of which 68 are protein-coding and 113 are either putative or noncoding, and there are 72 genes, with seven of them encoding putative transcription factors, in the critical interval. No individuals with XECD have been identified outside the original family, making analysis more challenging. In an effort to discover the causal gene, we have exhausted numerous technologies (WES, WGS, ONT, and PacBio), however, the specific variant responsible for causing XECD has not been elucidated. In an effort to identify new gene candidates, we present preliminary data utilizing Bionano optical genome mapping with RNA-Seq data to identify new gene candidates for this challenging Mendelian condition.

RNA sequencing approaches enable tissue specific B and T cell gene expression and immune repertoire profiling

Chen Song, Evan Janzen, Tessa Devoe, Ariel Erijman, Gautam Naishadham, Matthew Campbell

1 New England Biolabs, Ipswich, MA 01982 USA

The adaptive immune system defends the human body from pathogens through the recognition ability of antigen receptors from B and T cells. Different organs play various immune function roles, and therefore present different B and T cell gene expression and immune repertoire profiles. Abnormality of these profiles may cause immunological diseases and cancer. Next generation sequencing technology has facilitated the understanding of gene expression and related cell functions, but in-depth understanding on B and T cell phenotype and clonotype profiling is an emerging need to fully take advantage of sequencing technology for diagnosis and prognosis.

Here, we present a study on utilizing RNA sequencing approaches on immune system organs for B cell and T cell specific gene expression and full-length immune repertoire profiling. Total RNA was extracted from various tissue samples including healthy donor Peripheral blood mononuclear cells (PBMC), lymph node, bone marrow, spleen, thyroid, thymus and colon, as well as diseased PBMC and colon samples. The RNA samples were then made into RNA-seq and Immune-seq libraries which were sequenced on Illumina NextSeq2000. Gene expression analysis and immune profiling from RNA-seq data was performed with edgeR and TRUST4. Immune repertoire sequencing data was processed using pRESTO and IgBlast. The Correlation between RNA-seq and Immune-seq on BCR and TCR clonotype was analyzed. Differential expression and clonotype profiling comparison were performed between normal and diseased RNA samples.

This study has shown tissue specific B/T cell phenotype and clonality patterns that relate to tissue functions and disease progression. RNA-seq enables high multiplex immunophenotyping which is traditionally performed by flow cytometry. RNA-seq can detect high abundant clones that correlate with Immune-seq, while the latter can provide more in depth clonality signatures. Integrating both RNA sequencing and immune repertoire sequencing approaches allows accurate phenotype and clonotype determination for immune landscape in various tissues.

Seeing more of the brain with high-plex multi-omic imaging at single-cell spatial resolution in a large contiguous tissue section

Dwayne Dunaway[1], Zachary R. Lewis[1], Victoria Rachleff[2], Tushar Rane[1], Tiên Phan-Everson[1], Alyssa Rosenbloom[1], Shilah A. Bonnett[1], Giang Ong[1], Lidan Wu[1], Ashley Heck[1], Rustem Khafizov[1], David Ruff[1], Erin Piazza[1], Patrick Danaher[1], Aster Wardhani[1], Mithra Korukonda[1], Carl Brown[1], Christine Kang[1], John Lyssand[1], Jim Forrester[1], Trieu My Van[1], Vik Devgan[1], Jaemyeong Jung[1], Gary Geiss[1], Joseph M. Beechem[1], and C. Dirk Keene[2]

1 NanoString® Technologies, Seattle, WA, USA
2 University of Washington, Department of Laboratory Medicine and Pathology, Seattle, WA, USA

Neurophysiological and neurodegenerative processes occur over multiple spatial scales in the brain. While fine-scaled spatial molecular analyses of human brain function have vastly advanced in the last decade, these datasets often lack context of how microscopic features fit within the larger picture of macroscopic neuroanatomy. In the brain, critical cell functions and cell-to-cell communications take place over relatively large interstitial distances and far from the cell somas. The ability to explore protein and RNA-driven activities at high resolution, within the spatial context across a large environment and passing through multiple microenvironments, allows a comprehensive picture of brain biology to emerge and is applicable to open questions regarding brain development, activity, aging, and neurodegeneration.

A Spatial Molecular Imager (SMI) efficiently handles highly multiplex analysis of FFPE and fresh-frozen tissues at > 68-plex on protein and > 6,000 plex on RNA. Typical CosMx SMI assays make use of standard glass microscope slides with 300 mm2 of imaging area. However, the size, complexity, and heterogeneity of the human brain from one area to another further drive the need to sample much larger tissue sections. Here we demonstrate a novel Larger Surface Area flow cell with > 1,600 mm2 of imageable area. Fluidic and imaging conditions were optimized to deliver consistent performance.

(continue to page 103)

(continued from page 102)

Combining the Larger Surface Area flow-cell imaging capabilities with the human-specific > 68-plex CosMx™ SMI Human Neuroscience protein panel (Neural Cell Profiling and Alzheimer's Pathology), we demonstrate spatial imaging of neural cell and disease-specific targets (APP, Amyloid Betas, Tau), and key post-translational modifications. Advanced neuronal segmentation algorithms allow for specific tracing and segmentation of single neurons, including axons, across hundreds of micrometers of space. Furthermore, it is possible to leverage high-plex protein data from a large area of tissue to identify smaller regions of interest for 6,000-plex RNA profiling and obtain a more comprehensive view of the neurobiological picture. Ultimately, the ability to image and evaluate 5-fold larger sections of human brain tissues with unparalleled spatial context and high-plex analyte contents allows the grander scale of disease anatomy and processes to be cataloged with exquisite precision and accuracy.

Using long-read sequencing for genomic and epigenomic analysis in patient-derived samples with somatic IDH1 mosaicism

Carolina Montano[1], Luke Morina[1], Corina Antonescu[2], Elizabeth Wohler[2], Nara Sobreira[2], Winston Timp[1]

1 Department of Biomedical Engineering, Whiting School of Engineering, The Johns Hopkins University, Baltimore, MD, USA
2 Institute of Genetic Medicine, School of Medicine, The Johns Hopkins University, Baltimore, MD, USA

Ollier disease (OD) and Maffucci syndrome (MS) are rare cancer susceptibility syndromes characterized by multiple enchondromas (Ollier disease) in combination with hemangiomas (Maffucci syndrome). OD/MS arises in childhood and causes significant morbidity from severe skeletal deformities, pain, and pathological fractures. Approximately 80% of patients with OD/MS have enchondromas that exhibit somatic gain-of-function (GoF) variants in IDH1 and IDH2 genes. These variants occur early in embryogenesis, leading to a mosaic pattern of disease distribution. IDH variants can induce global DNA hypermethylation and promote malignant transformation in patients with sporadic gliomas, leukemias, and chondrosarcomas. Because no such mechanisms have been shown in tumors from patients with OD/MS, our objective was to demonstrate how GoF variants in IDH1 drive genome-wide methylation changes.

We used Oxford Nanopore Technologies (ONT) long-read whole-genome sequencing to identify genetic and DNA methylation variation simultaneously. ONT can detect large-scale SVs inaccessible to short-read sequencing and DNA methylation changes without PCR amplification or incomplete bisulfite conversion bias. We used DNA extracted from enchondromas of five patients with OD/MS carrying the mosaic IDH1 p.Arg132Cys variant (allele frequencies ranging from 8.6-13% identified by NGS) and five control chondrocyte samples for comparison. From this data, we called SNVs with the haplotype-aware variant caller PEPPER-Margin-DeepVariant pipeline, which were subsequently used to phase the reads into different haplotypes. We used Sniffles2 to identify structural variants, applying Jasmine to look for variants common to at least three of our five samples. Structural variants are also known to alter gene regulation via IDH1-independent mechanisms when they overlap with transcription-factor binding sites or involve regulatory non-coding regions. We used the ONT methylation data to generate haplotype-specific methylation frequency information, then adapted BSmooth to identify differentially methylated regions (DMRs) between enchondroma and unaffected cartilage from control samples.

(continue to page 105)

(continued from page 105)

Finally, we correlated SNVs, SVs, and methylation at the single molecule level to understand the genomic and epigenomic heterogeneity in somatic mosaicism. We expect our study to improve our knowledge of the genetic and epigenetic landscape of OD/MS, opening new horizons for more efficacious targeted agents to treat cancer susceptibility syndromes.

JOIN US AT AGBT 2024 BONAIRE 2 SUITE

Presentation:

WEDNESDAY, FEB 7, 3-3:20 PM
PALMS BALLROOM 2 & 3

Don't miss Emily Leproust, CEO and co-founder of Twist Bioscience, speaking on the latest insights and innovations from both Twist and beyond!

Emily M. Leproust, Ph.D.
CEO AND CO-FOUNDER
OF TWIST BIOSCIENCE

Posters:

TUESDAY, FEB 6, 1:30-3:30 PM
POSTER #650

Twist Pan-Cancer Reference Standards V2: Enhanced Precision and Reduced Errors in ctDNA Analysis

Lydia Bonar, MS

WEDNESDAY, FEB 7, 4:45-6:10 PM
POSTER #551

Development of a high-throughput NGS library preparation workflow with normalization adapters and in-line barcode technology

Owen Smith, PhD

Scan to schedule a consultative meeting, scientist to scientist.
To learn more, visit **twistbioscience.com.**

Other Biology Omics

A universal 6,000-plex RNA panel to construct comprehensive single-cell spatial atlases across multiple FFPE human tissues

Other Biology Omics

POSTER 401

Michael Patrick[1], Shanshan He[1], David Kroeppler[1], Olivia Hoshing[1], Patrick Danaher[1], Jason Reeves1, Melanie Moon[1], Jim Forrester[1], Trieu My Van[1], Julian Preciado[1], Haiyan Zhai[1] and Joseph Beechem[1]

1 NanoString® Technologies, Seattle WA.

Understanding the single-cell atlas and molecular organization of single-cell spatially resolved tissue is crucial to uncover underlying organ development processes and disease mechanisms. A high-plex RNA panel that has high coverage and works universally on various tissue types provides a powerful tool for researchers to study distinct molecular characteristics at spatial single-cell resolution. We have developed a universal 6K Discovery RNA panel that covers broad biological areas of interest with special emphasis on oncology, immunology, and neuroscience. This panel generates a high number of transcripts-per-cell (often over 1000 transcripts/cell) in intact formalin-fixed paraffin-embedded (FFPE) human tissue from many tissue types.

In this study, we used the 6K Discovery RNA panel to characterize eight different FFPE tissue types, including brain, skin, lung, breast, liver, colon, pancreas, and kidney from humans. Four protein markers and DAPI were co-detected on the same tissue slide to identify the morphology in tissues as well as to improve the accuracy of cell segmentation. Also, tertiary analysis algorithms were developed for cell typing, co-localization of genes and ligands, cell-cell interaction, and pathway analysis.

Hundreds of millions of transcripts were simultaneously detected with a Spatial Molecular Imager (SMI) with high sensitivity and specificity on an FFPE tissue section with up to 1 cm2 scan area. Thousands of genes were detected above the limit of detection (LOD) across each tissue, with single-cell and subcellular resolution. We also constructed the methods to investigate sample-specific spatial neighborhoods, defined by cell types, cell states, nearly a full-reactome set of biological pathways, and over 160,000 ligand-receptor pairwise interactions in each tissue type. Finally, we created the cell type and spatial neighborhood atlas of eight tissue types.

Single-cell spatial measurements at 6,000-plex in a large viewing area on archival tissue with the CosMxTM SMI, coupled with comprehensive tertiary analysis workflow, help researchers in every field to gain a global perspective of spatial transcriptional landscape across multiple FFPE tissue types, enabling the next level of biological discovery and translational research.

Application of genomic dose-response modeling to risk-based prioritization of contaminants detected in tributaries of the North American Great Lakes.

Jenna E. Cavallin[1], Kendra Bush[2], Kevin M. Flynn[1], Monique Hazemi[2], Erin Maloney[3], Daniel L. Villeneuve[1]

1 United States Environmental Protection Agency, Great Lakes Toxicology and Ecology Division, Duluth, MN, USA
2 Oak Ridge Institute for Science and Education, Research Participant, at US EPA Great Lakes Toxicology and Ecology Division, Duluth, MN, USA
3 University of Minnesota, Duluth, Department of Biology, Duluth, MN, USA

The United States Environmental Protection Agency (US EPA) and other international regulatory bodies have begun exploring the use of benchmark doses derived from genomic dose-response modeling for use in chemical screening, prioritization, and risk assessment. As part of this effort, genomic dose-response modeling was applied to facilitate a risk-based prioritization of contaminants detected in Great Lakes tributaries for which no previous ecological toxicity data were available. Larval fathead minnows (Pimephales promelas), the crustacean Daphnia magna, and the green algae Raphidocelis subcapitata were exposed seven concentrations of twelve different test chemicals for 24 h in a 96-well plate format. Homogenate samples were prepared and then analyzed using TempOSeq technology, the winner of a federal government challenge (TARGET EcoTox Challenge) aimed at facilitating a new paradigm in high throughput, genomics-based, ecotoxicology testing. Biologically active concentrations of the test chemicals, based on reduced genome concentration-response modeling, were calculated according to guidance developed by the National Institutes of Environmental Health Sciences National Toxicology Program. Biologically active concentrations of the tested contaminants ranged from 0.18 to 10.8 µM for fathead minnows, with fipronil carboxamide being the most potent. For Daphnia magna, biologically active concentrations ranged from 0.002 to 29 µM, with 3,4-dichlorophenyl isocyanate (3,4-DCI) identified as most potent. With the exception of 3,4-DCI, concentrations of the test chemicals detected in Great Lakes tributaries were at least 100-fold lower than those that elicited molecular responses in the exposed organisms, suggesting a low probability for adverse ecological effects. However, in the case of 3,4-DCI, detected surface water concentrations overlapped those that elicited a concerted molecular response in exposed Daphnia magna. Consequently, 3,4-DCI has been prioritized for additional toxicity testing and site-specific effects characterization. Results illustrate emerging applications of genomic concentration-response modeling for environmental protection. Contents of this abstract neither constitute nor necessarily reflect US EPA policy. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Biological, ecological, and genetic drivers of the fish mucosal microbiome revealed through multi-omic technologies

Jeremiah J Minich[1]*, Todd P Michael[1v]

*presenting author
1 Salk Institute for Biological Studies
Plant Molecular and Cellular Biology Laboratory

Fish have evolved to live in some of the most extreme environments, collectively contain more species than birds, mammals, reptiles, and amphibians combined, and are a critical source of protein for the world. One of the most remarkable features of fishes is that they are exposed to orders of magnitude more microbes than terrestrial vertebrates, thus are perhaps the most fascinating group of organisms to study in the context of the host-microbiome interactions. We collected and analyzed the microbiomes (gill, skin, midgut, and hindgut) from 101 species of marine fishes from Southern California spanning 22 orders, 55 families, and 83 genera (representing ~25% of local marine fish diversity) to investigate how the ecology, fish physiology, biology, and phylogeny impacts microbiomes. Host genetic similarity was associated with the microbiome in the gill, skin, and hindgut, but not midgut. We developed a novel method to estimate microbial biomass using sequencing data and found that gill microbial biomass was lower in larger, offshore, pelagic fish species. For three years (2017-2019), we sampled 230 fish from a fixed location (SIO pier, La Jolla, CA) to investigate the seasonal impacts on the mucosal microbiome of a single coastal pelagic species, Pacific mackerel. The gut microbiome did not differ seasonally while a strong seasonal pattern was observed in the gill and lesser extent the skin. Contrary to our hypothesis, this variation was most prominent between spring (high diversity) and fall (low diversity) corresponding to nutrient cycling rather than summer and winter (temperature). Within mackerel gills, Shewanella were the most dominant, prevalent (100% of samples), and was negatively associated with microbial diversity. Through respirometry experiments, we found that gill microbial biomass had no impact on swim performance (MMR) of S. japonicus. Next, we generated 8 high quality reference genomes with Oxford Nanopore Technologies (ONT) and Pacific Bioscience (PacBio) sequencing of various pelagic fishes (yellowtail, dorado, jack mackerel, Pacific sardine, Pacific mackerel, bonito, wahoo, Pacific bluefin tuna) to understand how pelagic fishes may have evolved to cope with high microbial loads in the gills as a function of changing environments.

Building the world's largest phage-host interaction atlas using proximity ligation technology

Benjamin Auch[1], Jonas Grove[1], Sam Bryson[1], Yunha Hwang[2], Demi Glidden[1], Emily Reister[1], Peter Girguis[2], Alexander Khoruts[3], Ivan Liachko[1]

1 Phase Genomics, Inc.
2 Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, USA
3 Division of Gastroenterology, Department of Medicine, Center for Immunology University of Minnesota, Minneapolis, MN, USA

Viruses, including bacteriophage and archaeal viruses, are the most abundant form of life on earth (1031). They interact with all life and shape the global ecosystem through their impacts on community composition and horizontal gene transfer. However, phage-host relationships have proven challenging to identify without use of culture-based experiments to generate unambiguous evidence for a phage's presence in a given host. These experiments inherently require that all hosts are culturable, typically restricting the scope and microbial diversity that can be surveyed and limiting our understanding of potentially valuable phage-host relationships.

Proximity ligation sequencing is a powerful genomic method for associating viruses with their hosts directly in native microbial communities. Proximity ligation captures, in vivo, physical interactions between the host microbial genome and the genetic material of both lytic and lysogenic phage. Similar to culturing experiments, these linkages offer direct evidence that phage sequences were present within an intact host cell, thereby establishing a phage-host pair. However, unlike culturing experiments, proximity-ligation methods do not require the propagation of living bacterial cells and unlike single cell sequencing experiments, only capture phage-host interactions inside cells. The combination of intra-phage and phage-host signal enables us to simultaneously deconvolve viral genome bins (vMAGs) directly from metagenomes and to assign microbial hosts to large numbers of vMAGs without culturing.

Our application of this technology to hundreds of complex microbiome samples has yielded thousands of novel phage and archaeal virus genomes with host assignments, as well as large numbers of new microbial genomes. Through broad-scale application of proximity ligation sequencing, we are creating a global-scale database of highly diverse phage-host interactions from samples from across the world. We will present published and unpublished work highlighting the power of this approach in the field of metagenomic discovery.

Dynamic transcriptome in gastrointestinal tracts at different lactation stages of dairy cattle

George E. Liu[1], Yahui Gao[1,2,3], Li Ma[2], Lingzhao Fang[4], Cong-jun Li[1], and Ransom L Baldwin VI[1]

1 Animal Genomics and Improvement Laboratory, Beltsville Agricultural Research Center, Agricultural Research Service, United States Department of Agriculture, Beltsville, Maryland 20705, USA
2 Department of Animal and Avian Sciences, University of Maryland, College Park, Maryland 20742, USA
3 Guangdong Laboratory of Lingnan Modern Agriculture, National Engineering Research Center for Breeding Swine Industry, Guangdong Provincial Key Lab of Agro-Animal Genomics and Molecular Breeding, College of Animal Science, South China Agricultural University, Guangzhou 510642, China
4 Center for Quantitative Genetics and Genomics (QGG), Aarhus University, Aarhus, Denmark

Lactation in dairy cattle is a critical period that demands significant adjustments in the rumen and digestive system to meet the increased nutrient requirements for milk production. Understanding the biological functions of the epithelia of the the rumen and digestive tract tissues is crucial for improving milk-related traits. In this study, we utilized next-generation sequencing and transcriptomic profiling to investigate the molecular basis of adaptation in the gastrointestinal tracts (epithelia of the colon, duodenum, and rumen) of dairy cattle during dry and entire lactation periods. By capturing gene expression patterns at multiple time points, we facilitated direct comparisons within and among tissues during different lactation stages, including early and peak lactation. Our computational analyses encompassed global transcriptome profiling at different lactation stages, identification of stage-specific genes, time series clustering of RNA-seq data, co-expression gene network analysis, tissue-specific expression patterns, functional enrichment, and cell type deconvolution. Through this comprehensive approach, we gained profound insights into cattle lactation, unveiling tissue-specific characteristics of the gastrointestinal tract and shedding light on the intricate molecular adaptations for milk synthesis occurring during lactation.

Epigenetic characterization of E14 mouse embryonic stem cells across a differentiation time course.

V K Chaithanya Ponnaluri, Daniel J. Evanich, Laura Blum, Sagnik Sen, Vaishnavi Panchapakesa, Ariel Erijman, Matthew A. Campbell, Bradley W. Langhorst, Sriharsa Pradhan, Louise Williams

DNA methylation (5mC) and hydroxymethylation (5hmC) are epigenetic modifications that have an important role in chromatin organization and regulation of gene expression. Techniques like DNA dot blot, immunofluorescence imaging and qPCR were used to establish the presence and dynamics of 5mC and 5hmC levels during early mammalian embryonic developmental stages. Next generation sequencing based techniques like TAB-seq and hMeDIP provided additional context to the distribution of 5hmC in the mammalian genome. However, these methods are dependent on the use of bisulfite conversion or an antibody binding to distinguish 5hmC from other forms of cytosine.

We developed NEBNext Enzymatic 5hmC-seq (E5hmC-seq), an enzymatic approach to identify 5hmC at single base resolution. This method employs T4-BGT and APOBEC enzymes to protect 5hmC signal and convert 5mC and cytosine bases to thymine and uracil respectively. These conversions enable accurate discrimination of 5hmC from 5mC and cytosine using input DNA ranging from 0.1 ng to 200 ng.

E14 mouse embryonic stem cells are used in the epigenetics field as a model system to evaluate the distribution of 5mC and 5hmC in the genome. Embryonic cells were differentiated over 10 days into neuronal lineage cells. We used NEBNext Enzymatic Methyl-seq (EM-seq) and E5hmC-seq to evaluate the dynamics of 5mC and 5hmC distribution in differentiated E14 cells at single base resolution. Interestingly, 5hmC levels decreased whereas 5mC levels increased particularly during the first five days of differentiation. LC-MS/MS quantification of this same DNA mirrored the changes observed by sequencing. Subtractive analysis of EM-seq signal (5mC + 5hmC) with E5hmC-seq signal (5hmC only) was performed to identify cytosines that showed only 5mC modification. Further, we correlated these 5mC/5hmC dynamics with transcriptional regulation to understand the crosstalk between epigenetic changes and transcriptional output.

GenomeTrakr database and network updates.

M. Allard, R. Timme, M. Balkey, S. Cianci, E. Stevens, M. Hoffmann, G. Kastanis, J. Payne, A. Pightling, H. Rand, J. Pettengill, Y. Luo, N. Gonzalez-Escalona, D. Melka, P. Curry, Y. Chen, S. Tallent, and E. Brown.

U.S. Food and Drug Administration, College Park MD USA. Marc.allard@fda.hhs.gov

A network of federal, state, academic, and international laboratories has been using WGS data to rapidly characterize pathogens since 2012. Sequences from this GenomeTrakr network are available through the NCBI Pathogen Detection web portal. Public health agencies (FDA, CDC, and USDA-FSIS) collect and share data in real time. This rapidly growing database is regularly used in outbreak investigations. The GenomeTrakr database has demonstrated how a distributed network of desktop WGS sequencers can be used in concert with traditional epidemiology for source tracking of foodborne pathogens. This "open data" model allows greater transparency between federal/state agencies, industry, academic, and international collaborators. This database has continued to grow and diversify increasing in the last year to ~1,200,000 draft foodborne genomes and over 1.6 million total human pathogen genomes. This report documents the value that is gained from analyses of a million pathogen genomes. NCBI, currently is producing daily phylogenetic results for 80 bacterial pathogens including: Salmonella, Listeria, E. coli and Campylobacter for > 120,000 clusters of < 50 SNPs, plus pairwise SNP count differences for cluster members. Another NCBI product called AMRFinderPlus provides genotype calls of presence/absence for known antimicrobial resistance genes as well as genotypes related to virulence and stress. We discuss GalaxyTrakr for global distribution of our bioinformatic pipelines, Protocols.io for sharing methods, GenomeGraphR which presents foodborne pathogen WGS data and associated curated metadata in a user-friendly interface that allows users to query a variety of research questions such as, transmission sources and dynamics, and persistence of genotypes associated with contamination in the food supply and foodborne illness across time or space. The FDA laboratory flexible funding model has distributed funds to double the existing network of domestic partners and has distributed funding for SARS-CoV-2 wastewater special surveillance project associated with human food workers. Results demonstrate global benefits of having an open data model. Understanding root causes of foodborne contamination and assisting our academic, public health and industry partners to develop preventative controls to make food safer globally.

Genomic dose response modeling in ecosystem protetion – innovation to support regulatory decision-making

Daniel L. Villeneuve[1], Helen Poynton[2], Garrett Evensen[2], Leah Wehmas[3], Monique Hazemi[4], Logan Everett[5], Jonathan Haselman[1], Adam Biales[1], Michelle Le[4], Kendra Bush[4], Kevin Flynn[1]

1 United States Environmental Protection Agency, Great Lakes Toxicology and Ecology Division
2 University of Massachusetts, Boston, School for the Environment
3 United States Environmental Protection Agency, Chemical Characterization and Exposure Division
4 Oak Ridge Institute for Science and Education, Research Participant, at US EPA Great Lakes Toxicology and Ecology Division
5 United States Environmental Protection Agency, Biomolecular and Computational Toxicology Division

Despite revolutionary advances in sequencing technologies, uptake and use of genomic data in chemical safety assessment, environmental protection, and associated regulatory decision-making has been limited. However, recently the US Environmental Protection Agency (EPA), National Institute of Environmental Health Sciences National Toxicology Program, Health Canada, Environment and Climate Change Canada, the Organization for Economic Cooperation and Development, and other authorities have begun exploring the use of benchmark doses derived from genomic dose-response modeling for use in chemical screening, prioritization, and risk assessment. As part of this effort, the US EPA launched a federal government challenge (TARGET EcoTox Challenge), promoted at AGBT in 2020, to stimulate development and demonstration of technologies that could support a nascent high throughput, genomics-based, ecological testing paradigm. The competition was held March 2020-December 2021. Solver teams were provided reference samples from four aquatic species (a fish, a crustacean, an insect, and an algae) that had been exposed to eight reference contaminants of varying modes of action. Reference samples were analyzed, and results submitted along with a technology description document. A total of five solutions were received from three independent solver teams and evaluated based on pre-defined scoring criteria. All solutions were judged to be compatible with the needs of the program; however, based on average ranks considering relative strengths and weaknesses with regard to cost, genome coverage, and quality, a winning technology, TempO-SeqTM was identified. Leveraging results of this government challenge, proof of concept studies have been completed for over 50 chemicals tested in two to four aquatic species. This presentation reports on results of the TARGET EcoTox Challenge and provides an overview of initial applications of genomic dose-response modeling in ecosystem protection, including a risk-based screening of per- and polyfluoroalkyl substances (PFAS). Contents of this abstract neither constitute nor necessarily reflect US EPA policy. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Investigating the role of dynamic DNA methylation changes in the innate immune response

Other Biology Omics

POSTER 410

Chad Hogan[1], Samira Asgari[1]

1 Institute for Genomic Health, Icahn School of Medicine, New York, NY, USA

Background: Upon infection with a pathogen, the immune system initiates a highly orchestrated change in gene expression to trigger the innate immune response. Epigenetic changes play a major role in this process. However, the role of DNA methylation, a ubiquitous epigenetic mechanism, has been largely overlooked in immune regulation, primarily because the common paradigm in the field considers methylation as a temporal, but permanent change associated primarily with development and differentiation. Recent studies challenge this paradigm; DNA methylation can be dynamically regulated in response to environmental stimuli such as infections. Yet, there is still much to understand about the role of DNA methylation in gene expression and the innate immune response.

Methods: To address this gap, we produced longitudinal (4 and 24hs) DNA methylation data from peripheral blood mononuclear cells (PBMCs) stimulated with lipopolysaccharide (LPS) (N = 7 individuals and 40 samples). Following data analysis with SeSAMe (v1.18.4), we performed principal component analysis (PCA) using the top 50% most variable probes, followed by testing the association of LPS stimulation with the logit-transformed ratio of the methylated probes (M values, N > 700,000 probes). We performed the analysis separately for 4h and 24hs post-stimulation PBMCs. Functional mapping and pathway enrichment analysis were performed, focusing on results with nominal significance (p-values < 0.01) due to our small sample size.

Results: For LPS stimulation, 6309 probes at 4h and 5250 probes at 24hs had nominally significant differential methylation. These probes were mapped to transcription start sites (TSS), yielding 465 and 250 gene-associated probes for each time point. Pathway enrichment analysis using Reactome pathways identified 98 and 82 significantly enriched pathways respectively, including IL-12 signaling and gene expression regulation at 4 hours, and fatty acid metabolism at 24 hours.

Conclusion: Stimulation of human PBMCs with LPS leads to differential DNA methylation. By aligning the differentially methylated probes with known TSS, we gain deeper insights into the pathways influenced by pathogen exposure and their impact on the immune system. These findings challenge the conventional view of DNA methylation as a stable, long-term modification, underscoring its adaptable nature in immune regulation.

Investigation of the human meiosis mutation rate through deep sperm and blood sequencing

Jeffrey A Rosenfeld[1], J Sebastian Garcia-Medina[2], Delia Tomoiaga[2], Jacqueline Proszynski[2], Krista Ryon[2], Christopher E Mason[2,3,4]

1 Tonix Pharmaceuticals, 26 Main Street, Suite 101, Chatham, NJ 07928
2 Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY, USA
3 The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY, USA
4 The WorldQuant Initiative for Quantitative Prediction, Weill Cornell Medicine, New York, NY, USA

The human mutation rate per generation is estimated to be approximately ~ 1e-8. However, this figure only includes meiosis events that lead to the conception of a baby and a resulting live birth. The actual mutation rate for meiosis is probably much higher, but can only be investigated by sample a raw product of meiosis. We have performed deep sequencing using novel technologies of matched blood and sperm genomes of human donors. Using high depth and unprecedented high quality sequencing from the Pacific Biosciences Onso system and the Ultima Genomics UG 100 system, we have been able to look at the mutation rate of individual sperm. Moreover, we used 10x Genomics single-nucleus RNA-sequencing (snRNA-seq) on a set of 5 sperm donors, including some who are parents to Autistic children, to characterize the expressivity of these mutations and to examine variants associated with neurodevelopment.

We have found wide variability in the sequence of sperm from an individual, with an order of magnitude (10-18x) more mutations evident in sperm that the baseline human inter-generational rate. Moreover, our data showed significant differences in protamine mutation rates between the cohorts (2.2x10-16). While Illumina NGS is only able to variations down to a level of 1% or 0.1% minor allele frequency (MAF), and cannot detect variation between individual sperm, here we were able to profile sperm directly and find evidence of not only a higher mutation rate, but also unique insertions, deletions (indels), and structural variations (SVs) that are possible clonal outgrowths in sperm, but not present in donor blood. Overall, these results contribute to the emerging view of extensive mosaicism in normal and diseased individuals, and have implications for health monitoring and in vitro fertilization clinics.

Multi-scale control of recombination and diversity in the mammalian germline

Florencia Pratto, Gang Cheng, R. Daniel Camerini-Otero

Genetics and Biochemistry Branch, NIDDK, NIH

Homologous recombination is essential for proper chromosome segregation during meiosis. Meiotic recombination is initiated by the introduction of programmed double-strand breaks (DSBs) in discrete parts of the genome called hotspots and its distribution influences patterns of inheritance and genome evolution. In the past years, we have gained insights on the molecular determinants of meiotic recombination initiation in mammals but the processes leading up to it remain largely unexplored. DNA replication precedes cell division in mitosis and meiosis but technical challenges have precluded the study of meiotic replication in mammals. We have previously developed an approach of mapping origins of replication from tissue, inferring replication timing profiles from meiotic S-phase nuclei and finally integrating these experimental data in an in silico model of DNA replication.

In human males, we identified a germline-specific patterning of replication timing; subtelomeric regions of human chromosomes replicate consistently early. This highly distinct pattern mirrors the distribution of male-specific DSBs, crossovers (CO) and non-crossovers (NCO). To further investigate the relationship between replication and recombination, we mapped origins of replication and derived replication timing profiles from human spermatocytes. We show that meiotic replication in humans is slower than mitotic replication. These well resolved meiotic replication timing profiles allowed us to assess differences in recombination initiation and outcomes. Crossovers in humans have been shown to be enriched in earlier replicating parts of the genome, but we found that this relationship is dominated by the subtelomeric regions. Early replicating DNA outside of subtelomeric regions is depleted for COs suggesting that the timing of DNA replication and GC content composition is not sufficient to explain all differences in recombination.

These data support a layered model of control of recombination and diversity where it is possible that different recombination pathways are favored in different parts of the genome. Additional factors are likely to play a role, such as the targeted increase of DSBs by organizing the chromatin on shorter loops.

Progress in the ruminant telomere-to-telomere (RT2T) project

Timothy P.L. Smith[1*], Brenda M. Murdoch[2*], Stephanie D. McKay[3*], Benjamin D. Rosen[4*]

On behalf of the Ruminant T2T Consortium

1 USDA, ARS, U.S. Meat Animal Research Center, Clay Center, Nebraska, U.S.A.

2 University of Idaho, Animal, Veterinary and Food Sciences, Moscow, Idaho, U.S.A.

3 University of Missouri, College of Agriculture, Food & Natural Resources, Columbia, Missouri, U.S.A.

4 USDA, ARS, Animal Genomics and Improvement Laboratory, Beltsville, Maryland, U.S.A.

*Co-first authors

The T2T project to generate the complete human genome revealed new insights into the structure and function of the previously "invisible" parts of the genome including centromeres, tandem repeat arrays, and segmental duplications. The project involved a unique ad hoc effort involving many researchers and laboratories, serving as a model for collaborative open science. Refinement of T2T processes now enables comparative analysis of complete genomes across entire evolutionary phylogenies to gain a broader understanding of the evolution of chromosome structure and function. The generation and analysis of T2T assemblies for multiple species represents a substantial increase in scale and would be daunting for any single laboratory. Efforts focused on the primate lineage continue to employ the successful open collaboration strategy and are revealing details of chromosomal evolution which may be general or lineage-specific features. The suborder Ruminantia has a rich history within the field of chromosome biology and includes a broad range of species at varying evolutionary distance with separation of tens of millions of years to subspecies still able to interbreed. An open collaborative effort dubbed the "Ruminant T2T Consortium" (RT2T) was formed to emulate the success of the human and primate initiatives in the Cetartiodactlya order, focusing on suborder Ruminantia. The initial RT2T assemblies of cattle, gaur, goat, bighorn sheep, and domestic sheep have been released in GenomeArk and initial observations from those assemblies and the status of the next seven species are described. We also present in detail the motivation, goals, and proposed comparative analyses of RT2T to examine chromosomal evolution in the context of natural selection and domestication of species for use as livestock and invite additional participants from the global research community.

Space research impact on life sciences: refining drug discovery via space medicine insights, boosting synthetic biology with space biology knowledge, and enhancing point-of-care solutions through innovative space technology

Daniela Bezdan[1,2,4], Kevin Achberger[3], Stefan Liebau[3], Daniel Kaschubek[4], Maria Birlem[4], Mark Kugel[4], Christian Bruderrek[4], Stephan Ossowski[1,2], Olaf Riess[1]

1 Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany
2 NGS Competence Center Tübingen (NCCT), University of Tübingen, Tübingen, Germany
3 Institute of Neuroanatomy & Developmental Biology (INDB), University of Tübingen, Tübingen, Germany
4 yuri GmbH, Meckenbeuren, Germany

The intersection of space exploration and life sciences is an emerging academic focus. With space biology influencing Earth's life sciences in recent years, here we present our initial findings.

Refining Drug Discovery via Space Medicine Insights
When astronauts travel to space, their bodies experience unique challenges, from radiation exposure to microgravity-induced physiological changes. Studying these effects is not merely to ensure astronaut safety; it provides a deeper understanding of human biology. Here we present an example of how we can use this knowledge to define early-onset targets in cancer.

Boosting Synthetic Biology with Space Biology Knowledge
Space presents an extreme environment where only certain microorganisms can survive. By studying these extremophiles, scientists can uncover novel metabolic pathways and biological mechanisms that can be co-opted for synthetic biology applications. For example, bacteria that thrive in space's harsh conditions might possess unique enzymes that can be utilized in biotechnological processes here on Earth. We collected and analyzed strains from three space missions from 2022-2023 not only to understand the capabilities of bacteria adapting to extreme outer worlds

Space technology is refining point-of-care approaches on Earth.
Automation efficiency is paramount in space missions and terrestrial healthcare. Yuri's ScienceShells, crafted in Germany and equipped with microfluid systems and automated microscopes, facilitate a broad spectrum of biomedical studies. We are gearing up to deploy 38 of these units to the space station in 2024. Here we present our partnership with the University of Tübingen to enhance the development of accessible, practical non-invasive health monitoring technologies, e.g. hair personalized medicine, bridging space, and Earth applications.

The urobiome and vaginal microbiome in women and correlations in urogenital tract wellness

Kelly Moffat[1], Baris Ozdinc[1], Manoj Dadlani[1]

1 CosmosID

Bladder and urinary microbiome (urobiome) communities have been characterized over the past several years both through culture and metagenomics-based methods. In this work we first evaluate the urinary tract microbiomes of women and identify microbial communities associated in healthy (N = 47) and dysbiotic (N = 43) states. The whole genome shotgun sequencing data was imported from publicly available studies and compared using taxonomic and functional analysis by CosmosID. We found that in addition to increased levels of pathogens in the dysbiotic samples, there are commensal microbial abundance signatures between the groups as well. Functional profiles differ significantly between these groups, driven by an increase in sporulation pathways in the dysbiotic group. Additionally, we evaluated urobiome and vaginal microbiome samples from the same women (N = 13). The vaginal microbiome is characterized by the dominance of Lactobacillus species, with some community state types showing more susceptibility to urogenital infections (for example, bacterial vaginosis and urinary tract infections). We found that there are functional and compositional similarities in bacteria, phages, and AMR genes between vaginal and urinary samples from the same women. As we understand more about microbial interactions and transitions between the urinary tract and vagina in women, we hope to help better understand and address many urogenital conditions and infections. These findings challenge conventional views of UTIs as solely pathogen-driven and raise the possibility of the urinary microbiome influencing susceptibility to infections and disease outcomes.

Unlocking spatially resolved transcriptomic and proteomic secrets of century-old lungs from the 1918 'spanish' influenza pandemic

Kirsty R Short[1*], Lauren E Steele[1,] Keng Yih Chew[1], James Monkman[2], Chin Wee Tan[2,3,4], Sebastien Calvignac-Spencer[5,6], Thomas Schnalke[7], Navena Widulin[7], Habib Sadeghirad[2], Alyssa Rosenbloom[8], Shilah Bonnett[8], Mark Conner[8], Murat Kekic[9], Eduard Winter[10], Joseph Beecham[8], Arutha Kulasinghe[2*]

1 School of Chemistry and Molecular Biosciences, University of Queensland, Brisbane, QLD, Australia
2 Frazer Institute, Faculty of Medicine, The University of Queensland, Brisbane, QLD, Australia
3 Division of Bioinformatics, Walter and Eliza Hall Institute of Medical Research, Melbourne, VIC, Australia
4 Department of Medical Biology, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Parkville, VIC, Australia
5 Department of Pathogen Evolution, Helmholtz Institute for One Health, Greifswald, Germany
6 Faculty of Mathematics and Natural Sciences, University of Greifswald, Greifswald, Germany
7 Berlin Museum of Medical History at the Charité, Berlin, Germany
8 Nanostring Technologies, Inc, Seattle, WA, USA
9 The Ainsworth Interactive Collection of Medical Pathology, The University of Sydney, Sydney, NSW, Australia
10 Natural History Museum Vienna, Vienna, Austria

The 1918 H1N1 influenza A pandemic is considered the deadliest influenza pandemic in recorded history and resulted in an estimated 50-100 million deaths worldwide. A striking feature of this pandemic was an unusual incidence of severe disease in otherwise healthy young adults. This pandemic was followed by another influenza virus pandemic in 1957, although this caused significantly less mortality (1-4 million deaths). Interestingly, the marked mortality in young adults seen in 1918 was also observed during the 1957 pandemic, although on a much smaller scale. It is thought that the severe disease in the young observed in 1918 was the result of a 'cytokine storm' or an unregulated immune response. However, this hypothesis remains purely speculative as no study to date has examined the immune response of young adults from the 1918 virus compared to other influenza virus pandemics. This is largely the result of a paucity of viable patient samples available from these pandemics. Spatial transcriptomics has emerged in recent years as a powerful tool to dissect the host response to infection. However, the feasibility of using this technique on historical (>100 years old) samples remains to be determined. Here, we utilised unbiased cutting-edge spatially resolved whole transcriptome (18,000 genes) and spatial proteomic atlas (570-plex) by the Nanostring GeoMx Digital Spatial Profiler to profile the lung architecture from patients aged 15-30 years infected with the 1918 H1N1 and 1957 H2N2 pandemic viruses. We show spatially resolved tissue signatures associated across multiple areas of infection, pneumonia, and lung anatomies. In this study, we provide the first comprehensive insight to the role of immunopathology in young adults during influenza virus pandemics.

NEBNext UltraExpress™ DNA, FS DNA and RNA Library Prep Kits

Sometimes speed is required to set you apart from the pack. The NEBNext UltraExpress DNA, FS DNA and RNA Library Prep Kits have been carefully optimized for speed, while providing the high yields and high quality that you've come to rely on. Each kit has a single-condition protocol, with fixed universal adaptor concentration and number of PCR cycles, for the ultimate in streamlining. With fewer workflow and cleanup steps and automation-friendly transfer volumes, the kits were built for ease of use and automation compatibility. And, they do this all while reducing consumables waste, making your discoveries at the bench greener.

Let NEBNext UltraExpress speed up your library prep with:

- Faster workflows
  - DNA and FS DNA library prep: < 2 hours
  - RNA library prep: 3 hours (or a single-day workflow when including poly(A) enrichment or rRNA depletion)
- Fewer steps & less consumables waste
- Wide input range
  - DNA: 10 – 200 ng DNA (mechanically sheared)
  - FS DNA: 10 – 200 ng DNA (intact)
  - RNA: 25 – 250 ng total RNA
- A single protocol for all inputs
- Automation compatibility

> *"Protocols designed to save time usually only save 20-30 minutes. The NEBNext UltraExpress RNA Library Prep protocol saved us just over an hour in processing time. This is quite significant. We were especially impressed with the new clean-up approach that resulted in squeaky clean libraries."*
>
> -Director, Large Sequencing Core

Visit New England Biolabs in **Antigua Suite 2** to learn more.

NEW ENGLAND Biolabs®

Technology and Application Development

A droplet-based microfluidic platform combined with primary template-directed amplification (PTA) enables the analysis of clonal heterogeneity and mosaicism in leukemia at single-cell genome level.

Technology & Application Development

POSTER 501

Nan Zong[1], Kyle Hukari[2], Joseph Dahl[2], Victor Weigman[2], Jay West[2], Stavros Stravrakis[1] and Andrew deMello[1]

1 ETH Zurich, Department of Chemistry and Applied Biosciences, Zurich, Zurich, Switzerland
2 BioSkryb Genomics, Durham, North Carolina, USA

Mosaicism in leukemia is driven by dynamic genetic instability within a heterogeneous population of cells, which introduces complexity into both diagnosis and treatment. The primary challenge is the detection of rare clones within a tumor. A secondary hurdle lies in the limitation of sequencing a substantial number of single cells to detect point mutations in rare clones. To address the first challenge, we leverage droplet microfluidics and Primary Template-directed Amplification (PTA) to develop a platform for whole genome amplification. Specifically, we encapsulate single cells within hydrogel beads to allow amplification of each cell's discrete genome. We demonstrate this capability using a patient-derived chronic myelogenous leukemia line, K-562, known for its genomic instability. Successful encapsulation of single cells within hydrogel beads allows for the formation of up to 1 million single-cell hydrogel beads/hour. These hydrogel beads were re-encapsulated with the PTA mixture within droplets for whole genome amplification, with real-time fluorescence monitoring of individual bead amplification. Detection of thousands of amplified single-cell genomes occurs in less than 1 hour. We then employ a microfluidic device to fragment genomes and attach unique barcodes to all fragments in the workflow. To address the secondary challenge, we develop a targeted approach for sample sequencing. First, we assess the performance of libraries at low sequencing depth to determine the presence of discrete barcodes and to ensure accurate insert structure. High-performing libraries are then sequenced to a greater depth to determine cellular mosaicism through copy number variation (CNV) at approximately 0.5X per cell. To determine single-nucleotide variant (SNV) status, clusters of cells with common aneuploidy structures are bioinformatically grouped to determine the most prevalent single nucleotide variants within sub-clonal populations. The accuracy of CNV-clustered SNV calling is verified with deeper whole genome sequencing of a subset of cells identified through low-pass sequencing. Data delineates the existence of single-cell CNV and SNV mosaicism at a scale that allows detection of rare emergent clones, which could be applied to clinical samples to determine disease evolution and define pathology mitigation strategies.

A long-read sequencing strategy for somatic structural variant detection in samples with low variant allele fractions

Maria E. Hernandez Gonzalez[1], Corinne Strawser[1], Saranga Wijeratne[1], Jenell Lewis[1], Kelli Roach[1], Huachun Zhong[1], Benjamin J. Kelly[1], Elaine R. Mardis[1,2,3], Richard K. Wilson[1,2], Katherine E. Miller[1,2], Tracy A. Bedrosian[1,2], Anthony R. Miller[1]

1 The Steve and Cindy Rasmussen Institute for Genomic Medicine, Abigail Wexner Research Institute at Nationwide Children's Hospital, 575 Children's Crossroad, Columbus, OH, 43215, USA
2 Department of Pediatrics, The Ohio State University College of Medicine, Columbus, OH, USA
3 Department of Neurosurgery, The Ohio State University College of Medicine, Columbus, OH, USA

Brain somatic mosaicism is a major cause of malformations of cortical development (MCD) and drug-resistant focal epilepsy in children. Causal somatic variants, which include structural variants (SVs), are identified in about half of patients. Therefore, identifying somatic SVs through genomic characterization is crucial for better understanding of disease pathogenesis. Several challenges remain for detecting SVs, which are even more apparent when SVs are present at mosaic (low) variant allele fractions (VAFs). SVs can be identified through short-read sequencing; however, short-read technologies all have a limitation in resolving SVs that are larger than a typical read length (150 – 300 bp). Long-read sequencing (up to 20 kb) improves the identification of SVs but provides lower read-depth impacting detection of low VAF events.

To increase read depth generated with long-read sequencing, we designed an approach that combines a hybridization-based target enrichment strategy, using a probe panel targeting 83 genes associated with focal epilepsy, with long-read concatenation to optimize detecting somatic mosaic SVs that affect disease-relevant genes.

To demonstrate the functionality of our approach, we first used commercial reference genomic DNA (gDNA) and constructed a long-molecule (10 – 12kb) Illumina library containing adapters with unique molecular identifiers (UMI). Next, we performed probe-based target enrichment with enriched molecules (5 – 6 kb in length) used as input into parallel PCR reactions to amend sequences needed for facilitating the concatenation process. In a subsequent ligation reaction, the final concatemer molecules were formed ("trimers"; three molecules ligated together). The concatenated molecules were used as input into PacBio library prep and sequenced on a single 8M SMRTCell.

(continue to page 126)

(continued from page 125)

Analysis included UMI deduplication to calculate on-target gene coverage and comparing data to a non-concatemer enriched library, and to a standard PacBio HiFi WGS from the same gDNA source. We show that our approach improves coverage over enrichment only, and greatly improves targeted gene depth (>100-fold) over standard PacBio HiFi WGS. In summary, this sequencing approach increases long read coverage per cost and may enable improved detection of low VAF somatic SV's, which can be applied to different disorders where low VAFs events are of interest such as focal epilepsy.

A mirror image sequencing by synthesis (SBS) chemistry that is biologically invisible - Design, synthesis and functional demonstration of (L)-form 3'-O-modified nucleotide reversible terminators with (D)-form engineered polymerase

Dae H. Kim, Bo-Shun Huang, Paramesh Jangili, Jonnadula V. S. Krishna, Jennie Lee

Dxome Inc., Irvine, CA 92618, USA

DNA sequencing by synthesis (SBS) using reversible termination modified nucleotides has been the biochemical engine propelling the exponential growth of genome sequencing last two decades. Here, we report for the first time, design, synthesis, and functional demonstration of in-situ SBS using a (L)-form 3'-O-modified nucleotide reversible terminators paired with a (D)-form engineered polymerase, both enantiomers (i.e., mirror image) of their natural biologically active forms.

Four (L)-form nucleotides (A, C, G, and U), chirally inverted from natural D-form, were modified as fluorescent reversible terminators by using an azidomethyl chemical moiety as a 3'-OH capping group and fluorescent dyes attached through a cleavable linker. To demonstrate the complete chirally inverted versions of the substrate and enzyme system, we chemically synthesized using solid-phase peptide synthesis (SPPS) and native chemical ligation (NCL), a 89-kDa (775 amino acids) high-fidelity mirror image mutant 9oN DNA polymerase that enables for the first time, sequencing of a completely synthetic mirror-image (L)-DNA using SBS method. (L)-form DNA templates containing homopolymer regions were accurately and specifically sequenced even in the presence of more than 4 orders of magnitude more amount of (D)-form DNA by using this new class of fluorescent reversible terminators on a prototype flow cell and four-color epi-fluorescent microscope.

(L)-DNA oligonucleotides are known have reciprocal DNA properties to their natural (D)-DNA counterpart, including hybridization kinetics while being highly resistant to biodegradation. These unique properties have led to considerable amount of interest in application areas such as DNA based data storage/computing, drug labeling, and bio-barcoding. The main limitation in further growth in these applications has been the inability to accurately sequence the mirror-image DNAs. (L)-form SBS chemistry has now opened paths to many dynamic in-vivo and in-situ applications where ability to accurately synthesize and sequence the (L)-DNA barcode/tag while maintaining molecular stability and specificity in complex biological environment is critically important.

A new spatial multi-omic approach for biological discoveries in colonic diseased tissues using a comprehensive Immuno-Oncology protein panel

Technology Development

POSTER 504

Shilah Bonnett[1], Christine Kang[1], Alyssa Rosenbloom[1], Mark Conner[1], Erin Piazza[1], Brian Filanoski[1], Rhonda Meredith[1], Hye Son Yi[1], Lori Hamanishi[1], Eduardo Ignacio[1], Nadine Nelson[2], Michael Prater[2], Vik Devgan[1], Rudy Van Eijsden[1], Melanie Moon[1], Lesley Isgur[1], Terence Theisen[1], Margaret Hoang[1], Gary Geiss[1], and Joseph M. Beechem[1*]

1 NanoString® Technologies, Seattle, WA, USA
2 Abcam plc, Discovery Drive, Cambridge, UK

The advancement of spatially resolved, multiplex proteomic and transcriptomic technologies has revolutionized and redefined the approaches to complex biological questions pertaining to tissue heterogeneity, tumor microenvironments, cellular interactions, cellular diversity, and therapeutic response. While spatial transcriptomics has traditionally led the way in plex, multiple studies have demonstrated a poor correlation between RNA expression and protein abundance, owing to transcriptional and translational regulation, target turnover, and most critically, post-translational protein modifications. Therefore, a more holistic, ultra-high-plex, and high-throughput proteomic atlas approach becomes critical for the next phase of discovery biology. Here, we present a barrier-breaking, spatial proteomics panel that was designed to accelerate scientific discoveries.

A Digital Spatial Profiler platform is uniquely suited to support high-plex proteomics, allowing for the simultaneous analyses of proteins from discrete regions of interest (ROIs) in FFPE tissue sections while preserving spatial context. The assay relies upon abcam antibodies coupled to photocleavable DNA barcodes readout with NGS sequencing, allowing for theoretically unlimited plex. Here we introduce the Human Immuno-Oncology Proteome Atlas (IPA), a 570+ antibody-based proteomic discovery panel, compatible with immunohistochemistry on FFPE tissues with NGS readout. IPA is the highest-plex, most comprehensive, antibody-based multi-omic panel to date focusing on key areas of immuno-oncology, oncology, immunology, epigenetics, metabolism, cell death, and signaling pathway regulation.

(continue to page 130)

(continued from page 129)

Here we demonstrate the performance of IPA on various cell lines and tissue. Additionally, we show the power of IPA, using the spatial multi-omic assay along with the GeoMx® Whole Transcriptome Atlas (> 18,000 transcripts), a 40-plex custom antibody panel and microbiome-curated RNA custom spike-in (~220 transcripts) to evaluate 70 different colon disease samples across 5 pathologies including adenocarcinoma, hyperplasia, and chronic inflammation. This is the highest-plex multi-omic (~610-plex proteins and >18,220 genes) study ever implemented for spatial biology. When we compared the diseased tissue to normal tissue, we observed an upregulation of specific pathways associated with tumorigenesis and inflammation. Furthermore, we observed distinct differences in proteomic and transcriptomic landscape between pathologies. The cutting-edge, data-driven, expert-curated IPA panel is at the forefront of spatial proteomics, empowering the researcher for the acceleration of biological discoveries.

A Novel Engineered Transposase (TnT™) for Improved Performance in Genomic Applications

Technology Development

POSTER 505

David Bays[1], Jonathan Vroom[2], Kalayan Krishnamoorthy2, Aksiniya Petkova[2], Ashley Silvia[1], Rebecca Feeley[1], Zac Zwirko[1], Maura Costello[1], Jack T. Leonard[1], Gavin Rush[1], Mathew Miller[3], Joseph Mellor[1]

1 seqWell Inc, R&D, Beverly, MA, USA
2 Codexis, Protein Engineering, Redwood City, CA, USA
3 Codexis, Life Science Technologies and Applications, Redwood City, CA, USA

Transposase-based chemistry underlies a number of popular sequencing methods, including NGS library prep across a range of applications, and epigenetic assays such as ATAC-seq and CUT&TAG. These methods and technologies are largely based on hyperactive mutants derived from Tn5 transposase that were first described over 20 years ago for applications in molecular cloning.

Transposase technology has become a powerful tool that enables efficient and highly scalable sequencing assays, largely due to its capability for fragmentation and appending sequence tags to DNA in a single step ("tagmentation"). Nonetheless, "classic" Tn5 transposase has performance limitations in NGS library prep in terms of sequence-specific bias and molecular conversion of DNA into library molecules, when compared to other library generation modalities such as mechanical or endonuclease-based fragmentation and ligation.

In this study, we report the engineering and performance of a novel transposase enzyme using the CodeEvolver® platform, which optimizes base enzymatic properties that correspond to key utility requirements in a wide range of genomic assays. Using a multistage directed evolution program initiated with a repertoire of divergent naturally occurring enzymes, we developed a novel non-Tn5 transposase protein that significantly improves (a) the sequence-specific insertional bias preference, (b) the specific activity and (c) tolerance of the transposase enzyme for inhibitory buffer or reaction mixtures, relative to canonical Tn5 transposase.

We describe a comparative performance assessment of this novel transposase enzyme (TnT) relative to conventional Tn5 transposase. TnT demonstrates significant improvements to coverage uniformity, GC bias, and higher overall library complexity for diverse sequencing applications.

A novel generation of single cell platform bridging live cell imaging and single cell genomics at scale

Magali Soumillon[1], Amara Alexander[1], Duane Juang[1], Nicolas Chevrier[1,2]

1 Flexomics, Waltham, MA, USA

2 Pritzker School of Molecular Engineering, the University of Chicago, Chicago, IL, USA

Single cell genomics has become an essential tool for the analysis of complex cell populations across all fields of biology. However, mainstream single cell genomics technologies only provide a snapshot of a given cell's state at a given point and do not allow for dynamic functional analysis.

Here we present a novel high-throughput platform bridging genomic and functional cellular analysis at single cell level. Our approach combines the favorite tools of cell biologists, including live cell fluorescent imaging, long-term culturing and functional cell assays to single cell genomics readouts in a flexible and easy to use solution. Tens to hundreds of thousands of cells can be isolated, cultured, assayed in our proprietary picowell arrays and imaged on standard fluorescence microscopes. Once cells have been functionally characterized, genetic material is captured on our pre-loaded and spatially mapped optically encoded barcoded beads for downstream single cell genomics analysis. This unique barcoding system allows for the direct correlation of single cell genomics data to individually imaged cells, thus combining dynamic cell and genomic measurements at single cell level in a high-throughput manner.

To demonstrate the capabilities of our platform, we show how key functional assays such as cellular proliferation, cytokine or antibody secretion can be miniaturized on our platform, measured over time and correlated to genomics readouts. We describe how we leverage both short and long-read technologies to connect and correlate gene expression (scRNA-seq) but also isoform analysis such as full-length T-cell receptors and immunoglobulins, to the key phenotypic attributes collected by fluorescent imaging.

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.

A novel hydrogel particle technology for capture and concentration of cfDNA

Technology Development

POSTER 507

Lauren P. Saunders[1], Stephanie Barksdale[1], James Erickson[1], Chamodya Ruhunusiri[1], Jared Obermeyer[1], Kaitlyn Schiffer[1], Anurag Patnaik[1], and Ben Lepene[1]

1 Ceres Nanosciences, Manassas, VA, USA

Traditional DNA isolation kits use a crowding agent or alcohol to modify DNA conformation, thus facilitating non-specific DNA binding to carboxylated or silica beads. This technology is not optimal for cell-free DNA (cfDNA) because larger DNA fragments bind more strongly than the smaller cfDNA fragments. We have created a novel approach based on hydrogel particles (Nanotrap® particles) functionalized with an affinity bait that binds strongly to DNA. This results in very high affinity binding of small DNA fragments accompanied by size exclusion of large DNA fragments. The hydrogel particles have been integrated into an optimized process for isolation of cfDNA from 1-4 mL of plasma collected in EDTA tubes. This DNA isolation process results in a very high concentration of cfDNA with minimal genomic DNA contamination. Because of the selectivity of Nanotrap particles for small DNA fragments, blood only has to be spun a single time to be suitable for extraction, resulting in an easier workflow. Nanotrap particles are magnetically functionalized and methods are readily automated. We compare this cfDNA isolation method with common commercially available cfDNA extraction kits and examine the differences in yield, purity, and mutation detection.

A novel modular method for high throughput, multiplexed normalization of NGS libraries

Kristina Giorda[1], Clara Ross[1], Martin Ranik[1], Travis Sanders[1], David Gelagay[2], Ariele Hanek[1], Ross Wadsworth[2], Trishana Nundalall[2], Deelan Doolabh[2], Thomas Harrison[1], Brian Kudlow[1]

1 Watchmaker Genomics, Applications Development, Boulder, Colorado, USA
2 Watchmaker Genomics, Applications Development, Cape Town, South Africa

In NGS library preparation workflows, library quantification and normalization, in which equimolar amounts of libraries are pooled for sequencing, is used to optimize sequencing throughput. The process is often tedious and error prone especially for users processing many samples. Standard time consuming protocols require quantification of individual libraries and adjusting them to equimolar ratios using electrophoresis, fluorometric, and/or qPCR based methods. Other methods may require proprietary library amplification primers, and destroy residual libraries, precluding resequencing.

To address these needs, we developed a novel method for NGS library normalization. Our method begins with an NGS library, the concentration of which is variable and unknown, and results in the retention of a defined number of molecules of this library. A normalization reagent, which binds stoichiometrically to a predetermined number of library molecules is utilized, followed by a magnetic bead capture that allows simultaneous normalization of up to 96 libraries across a 20 fold range of concentrations.

Our normalization method can be integrated at various stages in an NGS workflow, based on user preferences, for both singleplex and multiplex normalization. For singleplex normalization, the normalization reagent is added to each sample and briefly incubated, followed by removal of excess library. Next, captured DNA is eluted off the beads. The final pool is created by combining equivalent volumes of all libraries. The multiplex method allows the pooling of libraries after the normalization reagent incubation, prior to library capture, resulting in a single bead-based capture and elution for the entire pool.

(continue to page 135)

(continued from page 134)

We compared these techniques against a quantitative method based on the Agilent TapeStation and assessed relative reads counts by sequencing. We confirmed that libraries normalized using our novel system exhibited demultiplexing read count variances comparable to a traditional quantitative library normalization method. Furthermore, the sequencing data quality and secondary analysis results of libraries normalized by all three methods were identical.

In conclusion, our NGS library normalization module offers a highly simplified, streamlined, and non-destructive method of library normalization that reduces sample processing time significantly.

A Novel Transposase-based Method for Multiplexed Long Read Library Construction

Technology & Application Development

POSTER 509

Zac Zwirko[1], David Bays[2], Maura Costello[3], Gavin Rush[4], Joseph Mellor[5]

1-5 seqWell Inc., Research and Development, Beverly, MA, USA

Long read sequencing technologies are unlocking new frontiers in genomic analysis. The recently released RevioTM instrument from Pacific Biosciences provides 8X higher throughput than the Sequel IITM enabling long read applications at unprecedented scale. Despite this increase in sequencing throughput, challenges in Pacific Biosciences' library construction protocols remain unresolved. Fragmentation is typically performed through low-throughput mechanical means, and multiple micrograms of genomic DNA are needed to make a high-quality library. Overall preparation costs are also substantially higher than short read procedures due to the large number of enzymes needed, fragmentation consumables, and potential inclusion of automated gel fractionation.

Here we present a novel strategy for multiplexed whole genome library construction using plate-based fragmentation and sample indexing with a modified version of Tn5 transposase. After Tn5 treatment, samples are pooled for downstream enzymatic steps greatly reducing cost and input requirements for each sample. To provide a framework for multiplexing of mammalian genomes, one RevioTM run can accommodate up to four flowcells each with 25M Zero-Mode-Waveguides (ZMWs) for a total of 100M ZMWs per sequencing setup. Although some ZMWs remain unloaded to prevent mixed templates, this still equates to approximately 8 human genomes covered at 30X in each run assuming average fragment lengths >10kb. To provide proof-of-concept for multiplexed long read library construction, we used transposition to index and fragment 8 replicates of Human Genomic DNA (NA12878, Coriell) to an average size >15kb using 500ng per sample. Replicates were then pooled for library construction and sequencing on the RevioTM to mean coverage of approximately 30X per sample. We believe multiplexed library construction is a much-needed advancement in sample preparation that will usher in a new generation of high-throughput long read sequencing applications.

A seamlessly integrated workflow for neural cell phenotyping using high-plex spatial proteomics.

Technology Development

POSTER 510

Alyssa Rosenbloom[1], Shilah A Bonnett[1], Mark Conner[1], Tiên Phan-Everson[1], Zachary Lewis[1], Giang Ong[1], Yan Liang[1], Emily Brown[1], Liuliu Pan[1], Aster Wardhani[1], Mithra Korukonda[1], Carl Brown[1], Dwayne Dunaway[1], Edward Zhao[1], Dan McGuire[1], Sangsoon Woo[1], Brian Filanoski[1], Rhonda Meredith[1], Kan Chantranuvatana[1], Brian Birditt[1], Hye Son Yi[1], Vik Devgan[1], Rudy Van Eijsden[1], Melanie Moon[1], Erin Piazza[1], Jason Reeves[1], John Lyssand[1], Michael Rhodes[1], Gary Geiss[1], Joseph M Beechem[1]

1 NanoString® Technologies, Seattle, Washington, USA

The brain is complex and heterogeneous where cell function and cell-to-cell communication are critical for rapid and accurate performance. The ability to explore protein-driven activities at high resolution within the spatial context of their immediate environment is critical to gaining comprehensive pictures of brain development, activity, aging, disease or dysfunction, and inflammatory responses. Many existing approaches for high-plex single-cell spatial proteomics face issues around simplicity, speed, scalability, and big data analysis.

Here, we present an integrated workflow that addresses key concerns around high-plex proteomics. Spatial Molecular Imager (SMI) and Spatial Informatics Platform (SIP) comprise an end-to-end workflow that efficiently handles highly multiplex protein analysis at plex sizes exceeding 68 targets. The SIP provides full analysis support, including built-in or fully customizable analyses for cell typing, ligand-receptor analysis, neighborhood analysis and differential expression. The SMI protein assays use oligonucleotide-conjugated antibodies, that are detected using universal, multi-analyte readout reagents. The Mouse Neuroscience protein panel (Neural Cell Typing and Alzheimer's Pathology) is optimized to comprehensively profile neural cell lineages across the brain as well as the progression of Alzheimer's disease. Approximately 80% of the antibodies making up both panels are cross-reactive with human antigens.

(continue to page 138)

(continued from page 137)

The SMI protein assay reagents were validated on the FFPE adult mouse brain, mouse embryo, and Alzheimer's positive human brain. We used the Mouse Neuroscience protein panel with the SMI to identify multiple neuronal subtypes, different reactive states of astrocytes and microglia, cell degeneration and proliferation. In a single-cell exploration of microglia, we noted distinct patterning of key immune targets based on their immediate microenvironment. Additionally, we evaluated the co-expression patterns and activation states of microglia within 200 µm from the amyloid plaque and tau tangles.

The CosMx™ SMI is a high-plex spatial multi-omics platform that enables the detection of >68 proteins at subcellular resolution. In combination with the high-plex Mouse Neuroscience protein panel, we present a flexible and scalable informatics platform (AtoMx™ SIP), a robust solution for comprehensive neural and disease phenotyping that captures the complexity of neuronal and glial cellular activity with full spatial context.

A streamlined EM-seq method for whole genome and reduced representation methylation detection

Vaishnavi Panchapakesa, V. K. Chaithanya Ponnaluri, Daniel J. Evanich, Ariel Erijman, Romualdas Vaisvila, Bradley W. Langhorst, Louise Williams

New England Biolabs, Ipswich, MA 01938, USA

DNA methylation is an important epigenetic regulator of gene expression. In the human genome cytosines are methylated in the CpG context and are often clustered in CpG rich regions associated gene regulation. Traditionally, sodium bisulfite conversion was used to distinguish 5-methylcytosines (5mC) and 5-hydroxymethylcytosines (5hmC) from cytosines. However, this method damages DNA and introduces significant sequencing bias. NEBNextÒ Enzymatic Methyl-seq (EM-seqÔ)is an enzymatic approach that minimizes DNA damage therefore enabling longer insert sizes, lower duplication rates and a more accurate quantification of methylation.

EM-seq has been optimized further. Here we show equivalency between EM-seq and EM-seqv2 data for DNA inputs ranging from 100 pg to 200 ng. The EM-seqv2 methylation metrics for 200 ng to 100 pg inputs are consistent and libraries show improved library yields, consistent global methylation and an even GC bias representation. The increased library yields using this enhanced, streamlined protocol subsequently reduces sequencing duplications and costs. Cost is an important consideration when performing methylome analysis. Genome level methylation needs high sequencing depth to attain meaningful coverage to accurately call significant methylation differences across samples. Reduced representation libraries, a commonly used method to investigate DNA methylation, relies on enrichment of CpGs within libraries digested using MspI. These CpGs are commonly found in CpG regulatory regions. Although lacking comprehensive CpG coverage, for many researchers this method provides enough cost effective CpG methylation coverage for their studies.

Reduced Representation EM-seq libraries (RREM-seq) and Reduced Representation Bisulfite libraries (RRBS) were prepared using 200 ng to 100 pg of genomic DNA. EM-seq adaptors were ligated to MspI digested DNA followed by conversion using either EM-seq or sodium bisulfite. Libraries were PCR amplified and sequenced on an Illumina NovaSeq platform. RREM-seq resulted in higher yields and reproducible results regardless of input amounts compared to RRBS libraries. RREM-seq libraries also identified a higher number of CpGs with even coverage facilitating accurate methylation calling. RREM-seq libraries have superior sequencing metrics resulting in robust methylation profiling. Focusing on a subset of the genome generates higher coverage DNA methylation data at a lower DNA sequencing cost.

Adaptation of MAS-Seq/PacBio and ONT long read technology to the ResolveOME™ multiomic workflow for single-cell transcript isoform and DNA amplicon interrogation

Technology & Application Development

POSTER 513

Martin A. Newman, Isai Salas-Gonzalez, Victor J. Weigman, Jon S. Zawistowski

BioSkryb Genomics, Durham, NC.

The ability to detect accurately detect genomic structural variation (SV) is a crucial spoke in deciphering the multifaceted mechanisms of oncogenesis occurring in concert. There is often ambiguity in calling SV with short-read technologies, at both DNA and RNA levels of resolution, which can begin to be mitigated by long-read sequencing. Ascertainment of differential transcript isoform utilization (DTU) in disease or drug-resistant states also stands to directly benefit from long read and alleviate short read ambiguity. Here we embarked on studies to contrast SV and DTU calling between short and long read sequencing, at the single-cell level, with ResolveDNA genomic amplification and ResolveOME joint genomic/transcriptomic profiling. We employed two long-read platforms, PacBio Revio and Oxford Nanopore Technologies (ONT) PromethION to report DNA- or RNA-level single-cell data using a cell line panel of diverse types with diverse states of copy number alteration. Transcriptomically, we devised a MAS-Seq strategy with a 16-plex of matched parental and drug-resistant triple-negative breast cancer or acute myeloid leukemia single cells as well as euploid HG001 cells to compare the full-length ResolveOME cDNA directly sequenced by the long-read platforms or fragmented for short-read readouts. Genomically, we enriched for long ResolveDNA or ResolveOME amplification products and either enzymatically fragmented these for short-read sequencing or directly sequenced using both long-read platforms. Dependent on the researcher's biological interest, we demonstrate here the feasibility of a forked ResolveDNA/OME workflow that presents a long-read option.

Advancing tissue and cell preservation solutions using a novel reagent, BD® OMICS-Guard Sample Preservation Buffer, that preserves cell surface epitopes and mRNA for up to 72 hours prior to single-cell applications

Technology & Application Development

POSTER 514

Cynthia Sakofsky[1], Samatha Vadrevu[1], Hye-Won Song[2], Samuel Shi[2], Zhiqi Zhang[2], Manish Thakran[1], Aruna Ayer[1]

1 BD Biosciences, 2350 Qume Drive, San Jose, CA 95131
2 BD Biosciences, 10975 Torreyana Rd, San Diego, CA 92121

High-quality genomics data from single-cell experiments require upholding high-quality cells with good viability and unaltered biological states. Challenges, however, often arise in maintaining cell integrity during cell preparation steps prior to single-cell workflows. For example, geographical distance from the site of cell or tissue collection to the site of single-cell processing may require overnight shipment, taking up to 3 days. Flow cytometry cell sorting, cell stimulation or other cell preparation needs can also be time consuming, prolonging workflow timelines and risking changes to the transcriptome and/or proteome of the cells. Here, we showcase a novel cell preservation solution, the BD® OMICS-Guard Sample Preservation Buffer, which allows cells and tissues to be stored for up to 72-hours at 4 oC, while reliably capturing the true biology of the cell. This unique reagent gently preserves mRNA and protein integrity without the use of traditional cross-linking and harsh fixatives and allows increased flexibility of single-cell isolation and processing while expanding options for study designs.

Our data show high correlation of whole transcriptome (WTA) gene expression in PBMCs preserved for 72-hours versus non-preserved samples from Day-0. In addition to WTA assays, mRNA integrity was also maintained in targeted gene and VDJ full-length assays, suggesting robust preservation of select transcripts readily interrogated in immune-cell profiling. Cell surface epitope preservation was high in both preserved PBMCs and mouse spleen. For PBMCs, there was a high correlation of AbSeq molecule detection as compared to Day-0 samples. Importantly, gene expression in stress-associated genes and protein expression of AbSeq activation markers showed no evidence of altered cell states. Additionally, no major differences existed between cell-type distributions of preserved versus non-preserved PBMCs, underpinning that all cell-types can be recovered. Finally, cell retention rates were high (80%) after preservation for 72-hours. Together these data illustrate the ability of the BD® OMICS-Guard Buffer to effectively preserve both mRNA and protein for up to 72 hours, enhancing the flexibility of single-cell experiments.

For Research-Use Only. Not for use in diagnostic or therapeutic procedures.

An automated approach to enable population scale cost-effective single-cell sequencing of PBMCs from whole blood utilising the 10x Genomics Chromium X controller

Lesley Shirley[1], Abby Crackett[1], Sarah Cooper[1], Cristina Cotobal Martin[1], Gosia Trynka[1], Carla Jones[1], Emma Davenport[1], Thomas Whiteley[1], Katy Taylor[1], Tom Preston[1], Nicola Corton1, May Hu[1], Stephen Watt[1], Vivek Iyer[1], Krista Kortelainen[1], Ian Johnston[1], and Carl Anderson[1]

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

The field of human genetics has identified tens of thousands of genome regions significantly associated with human complex diseases and traits. To translate these findings into actionable therapies the next major challenge is to identify the genes, cell types and pathways perturbed by variants.

To address this, The Wellcome Sanger Institute (WSI) is developing a robust automated workflow to generate single-cell RNA sequencing data across tens of thousands of Peripheral Blood Mononuclear Cell (PBMC) samples, including upfront PBMC isolation and banking from whole blood.

PBMCs are isolated from daily deliveries of fresh donor blood draws from a network of collection sites in the United Kingdom. Central to this is the development of a miniaturised 96-well workflow, which utilises the Hamilton platform and EasySep™ Direct Human PBMC Isolation Kit. Isolated PBMCs are subsequently counted and banked.

Crucially, the ability to isolate and freeze PBMCs has opened up the opportunity for WSI to access the 10x Genomics HT protocol on the Chromium X platform, enabling highly cost-effective single-cell sequencing. Novel automated workflows on the Hamilton STARlet platform and LIMS-embedded logic will drive the pooling of PBMC samples and loading of the HT chip. Typically, 100-160 samples will be defrosted, pooled and loaded per run.

Loading the chip in an automated manner enables the processing of large sample numbers to be standardised and accurately tracked, mitigating the risk of sample swaps. Following processing on the Chromium X controller sample pools are retrieved and 10X 5' GEX libraries are subsequently prepared on the Hamilton STAR platform.

(continue to page 143)

(continued from page 142)

The pipeline will provide the flexibility for samples to enter at multiple points in the workflow, accommodating fresh whole blood, banked PBMC's and cDNA, maximising the accessibility for the diverse range of sample types at WSI.

The modular design approach aims to future-proof the line by building in agility for the adoption of future changes in technology, chemistries and for rapid expansion to support further applications such as 3'GEX, BCR/TCR, and CITE-Seq without the need to re-engineer the end to end pipeline.

An evaluation of Illumina Complete Long Read Sequencing

Nicholas Redshaw[1], Michael Quail[1], Jyoti Nangalia[1], Nick Williams[1], Andrew Shaver[2], Andrew Slatter[2], Michael Prettyman[2], Scott Westenberger[2], Pernille Albertus[2], Ian Johnston[1]

1 Wellcome Sanger Institute, Hinxton, Cambridge, United Kingdom
2 Illumina Inc., 5200 Illumina Way, San Diego, CA 92122, USA

Illumina's Complete Long Read (CLR) technology uses standard short read sequencing with Novaseq 6000 or X series instruments to generate contiguous long read sequences. Using the 1-2 day CLR library preparation workflow, we were able to generate N50s of average 5-6kb and phase blocks of ~150-200kb from 50ng DNA input amounts. With this data, we obtained good, even sequencing coverage across highly repetitive regions of genes such as STRC, HBA2 and CYP2D6, which have previously been shown to be very difficult to resolve with uniquely mapped short reads. The enhanced coverage seen across regions of these genes enabled improvements in downstream analysis, including increased haplotype resolution and variant calling accuracy.

An innovative engineering platform generates novel polymerases to overcome limitations in genomics applications

Julie Walker[1], Bjarne Faurholm[1], Leo Karamanof[1], Dave Matten[1], Daniel Le Jeune[1,2], Meghan Oddy[1,2], Eric van der Walt[1]

Watchmaker Genomics, Boulder, CO USA1, University of Cape Town, Cape Town, South Africa2

The past decade has seen explosive growth in applications that rely on sensitive and accurate detection of nucleic acids across a wide range of fields. As these spaces become increasingly specialized, the need for highly functional and purpose-built enzymatic tools has grown. Many advances in the genomics applications space are built upon the engineering of novel enzymes which permit access to information that was previously impossible to capture. To address such bottlenecks, we developed a proprietary enzyme engineering platform that allows us to select enzyme variants with features that are highly desirable in genomics applications workflows.

Our platform enables engineering of a wide selection of DNA- and RNA-modifying enzymes that utilize large-scale, hypothesis-free libraries as selection inputs to both enhance enzyme activity and/or introduce novel activities. To test our platform, we carried out comprehensive screening of single mutations across the coding sequence of a B-family polymerase against multiple selective conditions. Here we present results of selections for novel uracil tolerating mutations, as well as salt tolerance and enhanced processivity for amplification of long targets, in a B-family DNA polymerase.

Selections targeting uracil tolerance identified all amino acid positions in the polymerase uracil binding pocket which directly contact uracil bases as targets to diminish template uracil binding activity. In contrast to previous engineering efforts, quantitative data generated from the selections reveal the relative contribution of any given substitution to uracil tolerance. Surprisingly, the selections also uncovered novel substitutions at positions structurally distal to the uracil binding pocket that confer uracil tolerance. This dataset demonstrates the power of our enzyme engineering platform to select for polymerases with activities on templates generally inhibitory to the native polymerase. In addition, when libraries were challenged with high salt selections, the platform identified hundreds of mutations. A subset of identified mutations were deeply characterized and found to enable more efficient amplification of long targets, showing promise for applications in the long-read sequencing space.

Given the success of identifying polymerase variants with enhanced activities under a variety of challenges, our next focus is to identify novel polymerase variants that enable next-generation methylation detection workflows.

Application of inexpensive, short read sequencing to Ancient Genomes and Perturb-Seq

Matthew Coole[1], Andrew Bernier[1], Tom Howd[1], Florian Oberstrass[2], Doron Lipson[2], Martin Sosa[2], Sophie Low[1], Tyson Clark[2], Brianna Weller[2], Ariel Jaimovich[2], Jack Stohlman[1], Devin Sobezenski[1], Swapan Mallick[3], Nadin Rohland[3], David Reich[3], Niall Lennon[1], Stacey Gabriel[1]

1 Genomics Platform, Broad Institute of MIT and Harvard, Cambridge, MA, USA
2 Ultima Genomics, Newark, CA, USA
3 Department of Genetics, Harvard Medical School Boston, MA, USA

Many research areas require access to inexpensive short read sequencing to expand into new directions, new and/or larger cohorts to power discovery, and drive new applications. Multiple product launches over the past year offer varying degrees of improvement in cost, quality and/or throughput. It is important to recognize that not all platforms map well onto all applications, so choosing which sequencing technology to invest in to develop a protocol for should be determined in an application-specific manner.

Two applications that map well onto low-cost, high throughput short read sequencing are sequencing of naturally fragmented DNA and counting applications. We have evaluated the Ultima Genomics UG100 sequencer for exemplar projects and demonstrate a cost savings on the order of 25% based on the unique features of the UG100 to offer shorter flows and faster run times.

First, we show that the UG100 sequencer can analyze ancient remains (from over ~4000 years ago) exhibiting low-quality, degraded, and contaminated human DNA. Sequencing of ancient DNA gives important insights into evolutionary biology and allows for measurement of genetic differences to modern humans. As these ancient DNA samples are highly degraded, their insert size ranges from 75-150bp with high levels of contamination (bacterial) allowing only 41-53bp aligning to the human genome reference sequence and, on average, about 60% of sequences mapping to the human genome reference sequence. Inherently short fragments make it feasible to shorten the sequencing runs (less "flows," per Ultima's technology) and achieve two sequencing runs per reagent cartridge, driving down costs and lowering hands on time in the lab.

(continue to page 147)

(continued from page 146)

Additionally, we are expanding Perturb-seq (CRISPR-based screens with single-cell RNA-sequencing readouts) applications. As with short read ancient DNA sequencing runs, we are able to leverage one sequencing reagent kit for every two Perturb-seq sequencing runs, making the UG100 an effective sequencing platform for genomic counting applications. Further reducing sequencing costs is especially important to genomic counting applications, which currently suffer from high sample preparation costs.

Are ONT long read human whole genomes ready for production-scale?

Beth Marosy[1], Michelle Kokosinski[1], Jessica Gearhart[1], Justin Paschall[1], Brian Craig[1], Michelle Mawhinney[1], David Mohr[1], Molly Sheridan[1], Phillip Witmer[1], Donna Muzny[2], Alan Scott[1], Jessica Hosea[3], Carolina Montano[3], Luke Morina[3], Quihui Li[3], Alison Klein[4], Mike Schatz[3], Winston Timp[3], Kimberly Doheny[1] and the **All of Us Research Program**

1 Johns Hopkins University, Department of Genetic Medicine, Baltimore, MD, USA
2 Baylor College of Medicine, Human Genome Sequencing Center, Houston, TX, USA
3 Johns Hopkins University, Department of Biomedical Engineering, Baltimore, MD, USA
4 Johns Hopkins University, Department of Oncology, Baltimore, MD, USA

Long read sequencing has been a steadily developing technology with several advantages over short read sequencing including resolving complex structural variants, expansion repeats and phasing of variants. Johns Hopkins Genomics and the Center for Inherited Disease Research (CIDR) provide high-throughput genomic services and statistical genetics consultation to investigators working to discover genes that contribute to disease. Implementation of Oxford Nanopore Technologies long read sequencing in a high throughput core facility has required standardization and optimization of protocols to achieve 30x coverage whole genomes with an N50 >25kb in a reproducible manner. Typically, long read sequencing protocols require new extractions to facilitate high molecular weight DNA and increased DNA inputs to ensure higher N50s. Studies which use previously extracted DNA require optimization of protocols in order to achieve similar N50s. Here we provide insight from implementing this technology utilizing biobanked DNA from the NIH All of Us Research Program (AoU). The AoU protocol, developed at Baylor HGSC and optimized for longer N50s at CIDR, utilizes 2-3 ug of input DNA, g-Tube shearing and size selection on the Blue Pippin prior to LSK114 library prep and sequencing on a single R10.4 flowcell with 3 loads over 72 hours. To date we have processed over 150 samples achieving an average of 32x coverage and N50 of 28kb. Challenges to production-scale work have included the time required for super-high-accuracy live basecalling as well as inconsistencies in total yield of R10.4 flowcells. A second high priority study is examining familial pancreatic cancer samples with an expected total of at least 400 long read WGS. A pilot using 11 cell line derived DNA samples (fresh extractions optimized for long reads) achieved an average of 58x coverage and N50s of 27.5kb. In order to achieve this higher depth, we adapted our methods by increasing DNA input to LSK114 library prep to 6 ug, creating 2 libraries per sample, and sequencing 2 R10.4 flowcells per sample. Currently, we are actively exploring adaptive sampling, to address clinical use for repeat expansion testing and elucidation of "unidentified" 2nd variants in individuals with cystic fibrosis, to inform treatment options.

Assessing the ability of HiFi genome sequencing for identifying methylation defects

Christian Gilissen[1], Veronique Ivashchenko[1], Ronny Derks[1], Juliet Hampstead, Amber den Ouden, Gelana Khazeeva, Simone van den Heuvel[1], Raoul Timmermans[1], Jordi Corominas Galbany[1], Tom Hofste[1], Helger Yntema[1], Lisenka Vissers[1], Alexander Hoischen[1,2].

1 Department of Human Genetics and Radboudumc Research Institute for Medical Innovation, Radboud University Medical Center, Nijmegen, the Netherlands

2 Department of Internal Medicine, and Radboud Expertise Center for Immunodeficiency and Autoinflammation and Radboud Center for Infectious Disease (RCI), Radboud University Medical Center, Nijmegen, the Netherlands

Introduction

Several genomic defects can lead to aberrant genomic imprinting, including chromosomal imbalances, uniparental disomy (UPD), primary epimutations, or genomic mutations of imprinted genes. Recent advances in long-read sequencing technology now make it feasible to sequence full human genomes. In addition to comprehensively detecting human genome variation these technologies can simultaneously detect DNA modifications such as 5mC methylation. In this study we wanted to assess the ability of PacBio HiFi sequencing to (1) robustly identify genome-wide methylation, (2) identify specific differentially methylated loci, and (3) identify genome-wide epigenetic signatures in comparison to other approaches.

Material and Methods

We compared a Genome in a Bottle sample (HG002) for which HiFi sequencing data was available to methylation calls from two other published whole genome methylation assays (Enzymatic Methyl-Seq, Accel-NGS Methyl-Seq) in the EpiQC study. In addition, we sequenced 8 trios on the Sequel IIe instruments and one sample with a known uniparental disomy on the Revio instrument to 30x coverage in order to assess the ability to detect specific differentially methylated regions. To identify episignatures we performed HiFi sequencing for 12 patients with a known pathogenic variant in the KMT2A gene, and two patients with a KMT2A variant of unknown significance (VUS) on the PacBio Revio instrument.

(continue to page 150)

(continued from page 149)

Results

The three whole-genome methylation sequencing assays on HG002 shared ~93% of all reported CpG positions. Moreover, we observed that PacBio LRS detected methylation in additional ~5% of CpG sites. Using a set of 41 imprinted differentially methylated regions (6 paternal, 35 maternal) from literature we found a concordant imprinting for 31/41 regions (76%) across 24 inhouse samples. Using an HMM model we also successfully identified X-inactivated CpG islands in females. Finally, we could partially reproduce two existing KMT2A episignatures in PacBio methylation data of 6 patients. We are currently constructing a new episignature based on the full set of 12 KMT2A cases.

Conclusion

PacBio long-read sequencing technology provides reliable haplotype-resolved methylation data across the whole genome which can be used to identify aberrant disease-associated methylation patterns in patients.

Assessment of a Novel Digital Fluidics Platform for Rapid Sample Processing from Whole Blood to Library

Technology & Application Development

POSTER 521

Gregory Nakashian[1], Nicole Arruda[2], Gabriella Davis[2], Michelle Cipicchio[1], Ally Day[1], Brendan Blumenstiel[1]

1 Broad Institute of MIT and Harvard, Broad Clinical Labs, Cambridge MA, USA
2 Volta Labs, Inc., Boston, MA, USA

Relative to genomic sequencing technologies, library construction methods have been slow to evolve. Broad Clinical Labs has evaluated a walk-away, end-to-end DNA extraction to library construction platform from Volta Labs, which employs digital fluidics technology as opposed to a traditional microtube-based approach to control sample chemistry. Reactions are managed as programmable droplets using electrical fields, magnetic manipulation, non-contact mixing, and temperature control. This unconventional system allows for reaction optimization otherwise impossible in standard tube-based chemistry methods. Benefits of this technology include flexible scaling, operational efficiency, fast turnaround times (by eliminating sample transfers, the need for user intervention, and extensive user training), and improved sustainability (significantly fewer tubes, plates, and pipette tips).

A Volta Labs prototype system was independently evaluated by Broad Clinical Labs for two PCR-free WGS workflows: long read sequencing of libraries prepared with the PacBio SMRTBell Prep Kit 3.0, and short read sequencing of Illumina libraries prepared using enzymatic fragmentation and ligation. Both library construction workflows utilized whole blood as the input. DNA was extracted using Volta's HMW DNA extraction application. DNA yield and purity were measured for all samples. For PacBio PCR-free whole genome library construction, library yield and HiFi quality metrics were evaluated after sequencing on a PacBio Sequel IIe instrument. For short read library construction, library yield, fragment size distribution, and all standard Picard metrics for 30x PCR-free whole genomes were evaluated after sequencing on an Illumina NovaSeq 6000 system. Precision and sensitivity for SNPs and indels were also calculated. All libraries for long and short read sequencing were produced from whole blood samples in 4 to 6 hours with approximately 1 hour hands-on time, a process that traditionally takes 1 - 2 days.

Quality and sequencing metrics for all of DNA extractions and libraries generated from the Volta Digital Fluidics system were found to be comparable to those generated by Broad Clinical Labs' standard high throughput liquid handlers. The data shows this novel liquid handling platform has the potential to provide the flexibility to enable rapid turnaround clinical diagnostic workflows without compromising quality.

Automated CUT&RUN Enables Scalable Multiomic Mapping of Chromatin Proteins and DNA Methylation

Keith E. Maier[1], Matthew R. Marunde[1], Vishnu U. Sunitha Kumary[1], Carolina P. Lin[1], Danielle N. Maryanski[1], Ellen N. Weinzapfel1, Karlie N. Fedder-Semmes1, Liz Albertorio-Saez[1], Dughan J. Ahimovic[2], Michael J. Bale[2], Juliana J. Lee[3], Bryan J. Venters[1], Michael-Christopher Keogh[1], Immunological Genome Consortium

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

Understanding cell state – whether in relation to differentiation, disease state, or drug response, is central to numerous fields of research. ATAC-seq and RNA-seq are common approaches for cell characterization. However, these fail to provide mechanistic insight and lack the granular detail required to fully understand regulatory processes governing gene expression.

Epigenomic features – such as histone post-translational modifications and chromatin-associated proteins – mark distinct genomic compartments (e.g., promoters, enhancers) and regulate chromatin structure, gene expression, and cell function. Mapping these features provides a rich context to study cell fate and has great potential for revealing the basic mechanisms of cell differentiation. Despite this, efforts to integrate epigenomics into large scale cell type profiling efforts have been hampered by the poor sensitivity, high background, and low throughput of traditional chromatin mapping technology (i.e., ChIP-seq).

Here, we present automated CUT&RUN (autoCUT&RUN), a high throughput assay for rapid, ultra-sensitive profiling of epigenomic features from FACS-isolated primary cells, tissues, and cultured cell lines. This workflow generates reliable profiles from <10,000 cells per reaction, and is supported by a rigorous optimization strategy, high-quality antibodies, and quantitative spike-in controls.

As part of a multi-site collaboration with the Immunological Genome Consortium, we used our autoCUT&RUN platform to build a comprehensive epigenomic database of mouse immune cells - composed of >1,500 epigenomic profiles from >100 different FACS-isolated primary immune cell types.

(continue to page 153)

(continued from page 152)

Further, in two different approaches we show how CUT&RUN can be integrated with novel DNA methylation profiling techniques to deliver 1) a cost-effective whole genome methylation profile, and 2) a true 'multiomic' view of the co-occurrence of chromatin features and DNA methylation, which can reveal multivalent cellular mechanisms and deconvolute sample heterogeneity.

These studies set the stage to leverage high throughput epigenomics for previously intractable applications, including cell identity fingerprinting, biomarker discovery, drug development, translational research, and more.

NIAID Award ID: 1R44AI167215-01A1
NHGRI Award ID: 1R44HG011875-01A1

Automated high and mid-throughput nanopore library preparation from clinical samples for detection of viruses and other pathogens using a combination of standard liquid handling systems with digital microfluidics

Technology & Application Development

POSTER 524

Spyros Oikonomopoulos[1,2], Julio Serna-Vazquez[1,2], Mina Lotfi Kelahrodi[1,2], Anne-Marie Roy[1,2], Ju-Ling Liu[1,2], Jiannis Ragoussis[1,2,3]

1 Victor Phillip Dahdaleh Institute of Genomic Medicine, McGill University, Montreal, QC, Canada
2 Department of Human Genetics, McGill University, Montreal, QC, Canada
3 Department of Bioengineering, McGill University, Montreal, QC, Canada

The Covid-19 pandemic showcases how important it is the rapid and efficient viral genome sequencing for early detection and tracking of high-risk virus variants that inform public health policies. Devices that are easy and flexible to use for viral genome sequencing enhance the readiness to respond to pandemics. The combination of library automation and nanopore sequencing represents a rapid, flexible and mobile solution with the additional advantage that it allows the deployment in locations that do not have substantial genomics infrastructure and highly qualified personnel.

The commercially available MIRO CANVAS system (INTEGRA Biosciences) has been proven to be a versatile digital microfluidics platform for automatic library preparation for long-read sequencing platforms. Assays for the Oxford Nanopore Technologies protocols are currently available on the MIRO CANVAS system, though they permit the library preparation of only one sample per run. To process multiple clinical samples in parallel, we used a modified protocol that enabled us to automate the nanopore library preparation for a pool of 4 to 96 or more barcoded SARS-CoV-2 clinical samples. The samples were processed using the ARTIC v5 protocol for SARS-CoV-2 whole genome sequencing using tiled amplicons of 400bp size approximately and tagged with a common sample specific barcode. The modified protocol is an adaptation of the “ONT Native Barcoding kit V14 SQK-NBD114 protocol” on the MIRO CANVAS system.

(continue to page 155)

(continued from page 154)

We present the integration of the MIRO CANVAS system downstream of our high-throughput liquid handling systems used in-house for a “SARS-CoV-2 high sample throughput routine surveillance pipeline”, as well as the corresponding data collected from the PromethION runs. We additionally showcase the processing and the corresponding nanopore data from small numbers of pooled SARS-CoV-2 clinical samples, an assay suitable for laboratories that are routinely processing a limited number of samples coming from individual outbreaks.

Here we demonstrate the first PCR-free multiplexing library preparation for long-read sequencing using digital microfluidics on the MIRO CANVAS system that saves more than 2h of hands-on time for the operator. This allows to streamline processes for the handling of amplicon sequencing for different pathogens, but also bacterial and human DNA sample long read sequencing.

Automated Hybridization Enrichment of Hartwig OncoAct Panel Using Volta Labs Digital Fluidics Platform

Ewart de Bruijn[1], Erik van Dijk[1], David Koetsier[1], Joris Schoonderwoerd[1], Saikumar Karyala[2], Maryke Appel[2], and Edwin Cuppen[1]

1 Hartwig Medical Foundation, Amsterdam, The Netherlands;
2 VoltaLabs, Boston, MA, United States

Hybridization capture is routinely used in life science research and diagnostics to enrich targets for next-generation sequencing. Hybridization capture protocols require precise and reproducible sample handling under strict temperature conditions, particularly throughout the laborious series of bead-based manipulations following probe hybridization. In addition to inter-operator variability, multiple transfers between tips and reaction vessels result in unavoidable temperature fluctuations, ultimately impacting assay specificity and sensitivity, reproducibility, and sequencing economy.

In this study, we investigated the potential benefits of automating post-hybridization capture and wash steps on the Volta Labs Digital Fluidics platform, which does not rely on pipette-based liquid handling, and offers unique features such as continuous temperature control while merging, mixing, moving, and separating liquids and beads. Pre-capture Illumina libraries were prepared from mixtures of COLO829 tumor cell line pairs and NA12878/HG0001 gDNA, using the KAPA HyperPrep Kit and IDT adapters (50 ng gDNA inputs, 12 cycles of pre-capture amplification). Target enrichment was performed in 8-plex pools using the KAPA HyperCap 3.3 workflow and a custom ~4 Mb capture panel (Hartwig OncoAct Panel) covering 750+ oncogenes. The overnight hybridization incubation was performed manually, after which samples were transferred to a Volta instrument for hands-free execution of the capture and wash steps. Post-capture amplification (10 cycles) and cleanup were performed manually according to the KAPA HyperCap protocol. Paired end (2 x 150 bp) sequencing was performed using an Illumina NovaSeq 6000 instrument (>50 Gb raw data). Data were analyzed using the Hartwig Medical Foundation open source data analysis pipeline for panel-based cancer diagnostics.

Initial data analysis indicates that library pools partially prepared on the Volta platform are comparable to those performed manually, in terms of coverage and calling of clinically relevant variants. Data analysis is ongoing and we will report on a comprehensive set of sequencing metrics and variant calling results. This proof-of-concept study paves the way for routine, push-button automation of hybridization capture workflows on the Volta Digital Fluidics platform.

Automated mechanical and enzymatic tissue dissociation for reduced bias in single-cell sequencing data

Julie Laliberte[1], Lixiazi He[1], Sven Aschenbroich[1], Loreen Andrick[1], Juan Vicente Lorenzo[1], Martina Mankarious[1], Yunyen Chiu[1]

1 Singleron Biotechnologies GmbH, R&D, Cologne, NRW, Germany

Tissue dissociation is essential to isolate single cells from complex tissues for various applications, such as single-cell sequencing and primary cell culture systems. However, most existing methods rely on either mechanical force or enzymatic digestion, which have limitations. We propose a novel method that combines mechanical force and enzymatic digestion in a synergistic manner. Our fully automated method uses a custom device that applies controlled force to the tissue samples while exposing them to a mild enzyme mixture. We validated a broad input range of 10-4000 mg of different tissues. For example, the technology enabled a reproducible output of over 60,000 single cells of high viability and quality from as little as 20 mg of mouse brain tissue. The combination of mechanical and enzymatic tissue dissociation results in the completion of the procedure within 15 minutes, while maintaining high cell viability (>80%) and complete dissociation. The device is programmable, ensuring flexibility with up to 50 programs and customizable parameters such as motor speed and number of strokes. Our method has been validated on several mouse and human tissues, including pancreas, skin, prostate, ovaries, lung, brain, liver, kidney, spleen, breast and diseased samples as cancer and dystrophies. Furthermore, we could also demonstrate the successful dissociation of various models as organoids resembling intestine, pancreas, and brain into viable single-cell suspensions. Our method produces higher yield and viability than mechanical force alone and higher purity than enzymatic digestion alone. It preserves cellular identity, transcriptome integrity, resulting in reduced bias in single-cell RNA-Seq data.

Automated xGEN™ NGS Hybridization Capture and wash steps on the Volta Labs Digital Fluidics Platform Improves consistency and reduces hands-on time

Technology & Application Development

POSTER 527

Nick Downey[1], Saikumar Karyala[2], Brittany Niccum[1], Maryke Appel[2]

1Integrated DNA Technologies, Coralville, IA, USA
2Volta Labs, Inc., Boston, MA, USA

Hybridization capture is the gold standard for target enrichment in next-generation sequencing due to its sensitivity and scalability. However, protocols are long and laborious and meticulous execution is needed to guarantee consistent results. Volta's novel Digital Fluidics Platform offers precise droplet manipulation and continuous temperature control throughout all steps that involve paramagnetic beads, thereby offering unique strategies for automating and optimizing hyb capture protocols.

As a first step toward automating IDT's xGen NGS Hybridization Capture workflow on Volta's novel Digital Fluidics platform, we have focused on the capture and wash steps following hybridization of 5'-biotinylated probes to regions of interest in a genomic library. Libraries were prepared from HG001 DNA, Oncospan reference standards, or gDNA extracted from blood/saliva samples using the xGen DNA Library EZ UNI Kit or cfDNA & FFPE DNA Library Prep Kit (for OncoSpan cfDNA/FFPE samples) and xGen Normalase™ UDI Primers. Hybridization was performed with the xGen Exome Hyb Panel v2 or xGen Pan-Cancer Hyb Panel. All steps were completed manually up to this point, after which capture and wash steps were performed manually, or using the VoltaHyb™ Application for IDT xGen Capture & Wash on a Volta instrument. Select libraries were normalized using the xGen Normalase module. Post-capture amplification was performed using standard protocols and enriched library pools were sequenced on an Illumina NextSeq 2000 system.

Comparison of key library quality and sequencing metrics indicated that libraries partially processed on the Volta platform were of comparable or higher quality than those processed manually, with less variation across key metrics (such as on-target rate and uniformity). Variant calling results for HG001 exome libraries and OncoSpan libraries enriched with the Pan Cancer panel showed a high degree of correlation with truth data sets. Real-life samples from DNA extracted manually or on the Volta platform were processed with 100% success rate and low variance between technical replicates.

Apart from improvements to performance and consistency, partial automation of the xGen workflow on the Volta platform also significantly reduced hands-on time and plastic waste.

AVITI chemistry supports high content cellular profiling and cell morphology analysis

Sinan Arslan[1], Daniel Honigfort[1], Mariam Dawood[1], Connor Thompson[1], Ramreddy Tippana[1], Sanchari Saha[1], Tyler Lopez[1], Zachary Burns[1], Pin Ren[1], Jennifer Wong[1], Da Zhao[1], Jerry Wu[1], Colin Brown-Greaves[1], Jean Kwon[1], Yu Liu[1], Adeline Huizhen Mah[1], Derek Fuller[1], Ryan Kelley[1], Michael Previte[1]

1 Element Biosciences, San Diego, CA, USA

Currently, researchers who need sensitive, reliable information about RNA/protein expression and localization, genomic sequence, and morphology at the subcellular level on one sample need to use multiple assays and often multiple instruments. Each assay can be time and resource consuming, susceptible to batch effects, and may have sample preparation requirements that are mutually incompatible, leading to difficult choices in experimental design. We present an advancement in the Element Biosciences AVITI™ sequencing platform that allows researchers to collapse multiomics data collection into a single assay by enabling simultaneous in situ sequencing and multiomic profiling of cells. This is achieved through Avidity Sequencing™, a novel probe scheme, subcellular resolution imaging, and integrated AI-based primary analysis for single-cell and spatial applications.

With a flexible, 30-minute hands-on time workflow, the AVITI platform can concurrently assess RNA and protein expression levels in situ at high plexity, capture detailed cell morphology data, and sequence targeted RNA regions within the same sample. In preliminary studies, we demonstrate these capabilities under specific biological conditions across different cell types. The platform can sensitively detect changes in mRNA in response to metabolic stress and cytokine regulation, showing changes of 60-fold over a dynamic range of 5 to >1000 copy number per cell, per gene. For protein detection, the AVITI antibody probes achieve >90% subcellular co-localization with data from benchtop methods such as immunohistochemistry while using significantly less reagents in comparison. The AVITI cell morphology method allows more features and organelles to be imaged simultaneously than a standard Cell Painting assay, on 1M to 4M cells per flow cell. Finally, we demonstrate direct sequencing of V(D)J segments in captured T lymphocytes on AVITI flow cells, leveraging the unique advantages of low dye concentrations in Avidity Sequencing, high resolution imaging, and low cycle times.

These assays are performed using a standard commercial AVITI system and cartridge with the addition of new probe chemistries for the cell readout. With this new functionality, the AVITI system can return data on mRNA, protein, targeted sequencing, cell paint phenotyping and morphology with one box, in one experiment, in one day.

Balanced-strand sequencing enables massively scalable duplex variant calling of ultra-low frequency mutations

Alexandre Pellan Cheng[1,2], Srinivas Rajagopalan[1,2], Majd Al Assaad[2], Aaron Sossin[1,2], Sunil Deochand[1,2], Adam Widman[1,3], Soren Germer[1], Zoe Steinsnyder[1], Dina Manaa[1], Melissa Marton[1], Catherine Reeves[1], Itai Rusinek[4], Eti Meiri[4], Omer Barad[4], Zohar Shipony[4], Shlomit Gilad[4], Ariel Jaimovich[4], Michael Sigouros[5], Jyothi Manohar[5], Abigail King[5], David Wilkes[5], John Otilano5, Olivier Elemento[5,6,7]; Bishoy M. Faltas[2,5,8,9], Juan Miguel Mosquera[5,6,7], Dan-Avi Landau[1,2]

1Integrated DNA Technologies, Coralville, IA, USA
2Volta Labs, Inc., Boston, MA, USA

Identifying single nucleotide variants (SNVs) is fundamental to genomics. Traditional sequencing-based variant discovery relies on consensus mutation calling requiring multiple variant-containing reads, which is challenging for ultra-low mutation frequency contexts, including circulating tumor DNA (ctDNA) or somatic mosaicism. Here, only one supporting read may be detectable and difficult to distinguish from common sources of variation such as sequencing error.

To suppress these errors and enable low frequency SNV detection, molecular barcoding and redundant sequencing have been proposed. Notably, duplex sequencing, where corresponding Watson and Crick strands are jointly analyzed for variants, can greatly reduce pre-analytical artifacts. Recent pioneering work has shown that duplex sequencing can decrease error rates to 10-7 (Hoang et al, 2016) and can be deployed genome-wide (Cheng et al, 2023). Nonetheless, this approach requires massive over-sequencing, as strand separation during DNA amplification challenges the ability to capture both strands. These inefficiencies result in use of only 1-5% of unique reads, limiting coverage depth.

(continue to page 161)

(continued from page 160)

We present balanced-strand sequencing, a PCR-free method that leverages the hydrogen bonds of complementary DNA to carry both native strands into sequencing. Balanced-strand sequencing is enabled by Ultima Genomics sequencing, where DNA denaturation is not required prior to clonal amplification. Specifically, the double-stranded DNA is partitioned and clonally amplified on sequencing beads through emulsion-PCR. Each double-stranded DNA molecule contributes to a single sequencing read, allowing for linear increase in duplex recovery with greater sequencing depth. Importantly, balanced-strand sequencing does not require redundant sequencing, thus maximizing unique molecule throughput. Integrating balanced-strand sequencing with machine-learning guided single read mutation calling reduced error rates to below 10-7, sufficient for rare variant discovery in tissue somatic mosaicism.

Applied to ctDNA, with as low as 5ng inputs, we achieved 7-35x whole-genome duplex coverage, ~20-fold higher than prior duplex WGS, with accurate ctDNA detection in the minimal residual disease context at concentrations <10-5. In early disease and post-therapy, we detected APOBEC and platinum exposure signatures to simultaneously track tumor-derived material and the effect of genotoxic treatment. Altogether, we envision that this technology will empower detection of rare variants and discovery of somatic mosaicism, translating to transformative impact on disease profiling.

Barcoding-free high-throughput single-cell targeted sequencing enabled by the 1-read-1-cell paradigm

Technology & Application Development

POSTER 530

A.Watson[1], V. Kurmauskaitė[2], V. Kiseliovas[2], R. Žilionis[2]

1 Atrandi Biosciences, Denver, Colorado, USA
2 Atrandi Biosciences | Droplet Genomics, Vilnius, Lithuania

High-throughput single-cell sequencing is largely based on the use of cell barcodes. However, for lower target numbers (2-100s) studied in large numbers of cells (10,000-1,000,000), the use of cellular barcodes might be excessive, as every cell needs to be presented with a unique set of DNA oligonucleotides. This is expensive and technically challenging, even when performed in microfluidic droplets. Here, we present the 1-read-1-cell principle for sequencing large numbers of cells in a barcoding-free fashion. Rather than being linked by barcode information, DNA targets originating from the same cell are concatenated into a single fragment addressable by long-read sequencing. Our Semi-Permeable Capsules (SPCs) allow the application of the 1-read-1-cell principle at high throughput. SPCs take advantage of the single-cell compartmentalization rates characteristic of droplet microfluidics. However, in contrast to regular water-in-oil droplets, SPCs enable straightforward multiple-step reactions without loss of compartmentalization.

For the proof-of-concept experiment, we used a mixture of E. coli and B. subtilis cells encapsulated into SPCs. We used two E. coli strains of which one strain was transformed with pUC19 plasmid containing AmpR and GFP-encoding genes, thus the final mixture of cells was expected to contain ~33% of B. subtilis cells and ~67% of E. coli cells (~33% transformed with pUC19). We amplified three different genes (16S rRNA, GFP-encoding gene, and AmpR-encoding gene), concatenated amplicons within SPCs, and sequenced the resulting ligation products on the Oxford Nanopore platform. Only plasmid-transformed E. coli cells were expected to produce long concatenated reads containing all three amplicons. Indeed, 92 % of ⩾ 3000 bp reads that satisfied quality parameters were E. coli 16S rRNA, AmpR, and GFP concatemers, while only 0.5 % were mixed-species reads. Enrichment of the 16S rRNA gene enabled the validation of the bacterial composition of the initial sample. In this experiment, 70 % of all reads were of E. coli origin and 25 % were of B. subtilis which is close to the theoretical prediction. We conclude that the 1-read-1-cell principle is an attractive tool for target linking on a single-cell level applicable even to hard-to-lyse microorganisms.

Beyond the Genome: Unraveling Protein Variability with Quantum-Si's Next-Generation Protein Sequencing Technology

Technology & Application Development

POSTER 531

John Vieceli, Khanh D. Q.Nguyen

1 Atrandi Biosciences, Denver, Colorado, USA
2 Atrandi Biosciences | Droplet Genomics, Vilnius, Lithuania

Protein sequencing is an invaluable tool that advances genomics research by providing crucial insights into the functional proteins encoded by the genome and offering a comprehensive understanding of cellular processes and disease mechanisms. In this study, we present how the next-generation protein sequencing Platinum™ platform allows us to probe protein variability at an unprecedented level of detail. In this approach, individual peptides are immobilized on a semiconductor chip and probed by dye-labeled N-terminal amino acid (NAA) recognizers followed by subsequent aminopeptidase cleavage exposing each NAA in the peptide for recognition. Fluorescent intensity, lifetime, and binding kinetics of each NAA binding event is recorded and converted into amino acid calls. This information-rich capability enables recognizers to discern single amino acid substitutions and in turn distinguish protein variants.

To demonstrate the versatility of this core methodology and its kinetic principles across a variety of proteins, we sequenced three variants of SARS-CoV-2 spike proteins: Alpha, Delta, and Omicron. Using Platinum, we successfully discriminated two single amino acid substitutions among these variants, effectively distinguishing them from one another. Additionally, Platinum was utilized to successfully discern a single amino acid substitution at the 12th position of a peptide in ubiquitin, showcasing the sensitivity of the system in the discovery of mutations deep into peptides. Finally, recent developments in our analysis algorithms have allowed us to sequence proteins without prior knowledge of their identities, enabling protein inferences tailored to specific biological pathways. As we envision future enhancements to our system, we anticipate an expanded coverage of the proteome, coupled with heightened precision in distinguishing protein variants down to the amino-acid level.

Bleed to read: From blood samples to ready-to-sequence NGS libraries.

Technology & Application Development

POSTER 532

Shreyas Shay[1], Anubhav Tripathi[2], Román Paz[3]

1Revvity, Applied Genomics, Hopkinton, MA, USA
2Brown University, Providence, RI, USA
3Revvity, Applied Genomics, Hopkinton, MA, USA

1 Atrandi Biosciences, Denver, Colorado, USA
2 Atrandi Biosciences | Droplet Genomics, Vilnius, Lithuania

Next-Generation Sequencing (NGS) is becoming readily accessible; however, nucleic acid purification and preparation of sequencing ready libraries still stand as bottlenecks in NGS protocols, particularly in low throughput workflows. Part of this is due to the lack of a reliable and user-friendly complete automation system for low throughput nucleic acid purification and library preparation. Revvity's BioQule™ NGS System can fully automate both the nucleic acid purification and library preparation for up to 8 samples in a single workflow, reducing the risk of human error and minimizing hands-on time to only 20 minutes to go from sample to quantitated NGS library. This fully walk-away sample preparation and quantification benchtop system comes at low-cost for automating library preparation directly from human whole blood.

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

Bridging the Genomic Divide: A Compiler-Centric Approach to Inclusive Genomic Analysis for Visual Impairment

God'salvation Oguibe

Advancements in genomics have revolutionized medical diagnostics and personalized treatment, yet challenges persist for individuals with visual disabilities to access the benefits of DNA analysis. In response to this critical need, a pioneering innovation is being proposed—a novel compiler-centric approach that enhances the speed and accessibility of DNA analysis for these visual impairment populations. This conceptual technology marries the power of computer science with the imperative of inclusivity and empowerment in healthcare. The innovation harnesses the principles of compiler design, traditionally used to convert human-readable code into machine-executable instructions. The compiler approach acts as an intermediary between raw genetic data and specialized software interfaces designed for the visually impaired by automatically translating complex genetic sequences into comprehensible, multimodal outputs, such as auditory and tactile representations, thereby democratizing the insights unlocked by DNA analysis. The compiler translates genetic information into audio descriptions, enabling them to independently interpret genetic reports. This idea represents a significant leap toward equitable access to genomic information. By accelerating the analysis process and providing accessible outputs, it empowers the visually impaired to engage more actively in their healthcare decisions. Moreover, it contributes to a more inclusive and compassionate healthcare system, where no one is left behind in the genomics revolution. The compiler approach demonstrates the potential for technology to close the gap between complex scientific data and its meaningful impact on the lives of those with diverse needs.

Keywords: DNA, Visual Impairment, Compiler, HealthCare

Broadening the input application of ResolveDNA genomic amplification to detection of SNV in circulating DNA

Jon S. Zawistowski, Swetha D. Velivela, Isai Salas-Gonzalez, Megan Justice, Tia A. Tate, Victor J. Weigman

BioSkryb Genomics, Durham, NC.

Uniform, allelically-balanced, and near-complete amplification of single-cell genomes has been enabled by ResolveDNA technology powered by primary template-directed amplification (PTA). This is achieved by attenuation of amplicon size, whereby the short amplicons have a low propensity to reamplify, manifesting with quasi-linear vs exponential amplification. Despite the power of this approach, this presents an inherent challenge for the amplification of cell-free and circulating tumor DNAs (cfDNA/ctDNA), which continue to hold promise as non-invasive diagnostic and monitoring tools for a multitude of disorders, due to their small size of only several hundred base pairs available as PTA input template. To provide a solution to the chief cf/ctDNA problem of template limitation, yet maintain the fidelity of PTA, we designed a scheme to concatenate short DNA fragments prior to amplification, increasing the size of input template to that which is amenable and optimal for PTA. This was undertaken using a size titration of synthetic DNA fragments as well as with cfDNA reference standards. Upon low-coverage sequencing, extending the template size of small fragments by ligation improved uniformity of exonic coverage as well as the number of loci detected within the cfDNA reference standard. Importantly, upon high-pass sequencing, while only 5/386 cfDNA reference standard loci were exposed with one detected single nucleotide variant (SNV) in the absence of our ligation strategy, 167 sites with 43 verified SNVs were detected post-ligation. Studies are underway to extend this proof of principle experiment to cfDNA isolated from plasma of a patient harboring a late-stage solid tumor with a defined missense mutational profile from a pan-cancer targeted panel, as well as to contrast the mutation detection sensitivity of existing methodology of cfDNA sequencing in the absence of amplification. In summary, the strategy presented here exposes single nucleotide variation from a realm of template previously not exposed with PTA, and holds potential for assessment not only for non-invasive detection of tumor mutations but for a myriad of additional cfDNA applications including prenatal screening and heart disease.

Building the spectrum of ground truth genetic variation in a four-generation 28-member CEPH family

Zev Kronenberg

Comprehensive ground truth data is required for validating sequencing pipelines in clinical settings, assessing the strengths and weaknesses of genome sequencing technologies, and improving variant detection software. Truth sets of genomic variation have lagged advances in sequencing accuracy and completeness; furthermore, current truth sets are confined to easy-to-characterize regions that comprise ~80% of the human genome. These benchmarking sets are mostly limited to small variants, missing complex variants like SVs and mega-bases of tandem repeat variation. Developing more complete truth sets will spur improvement in variant calling algorithms in under-characterized regions of the genome.

We have built a new truth set, leveraging the power of a large four generation 28-member family (CEPH 1463) using multiple sequencing technologies (PacBio HiFi and SBB, ONT, Illumina, and Strand-seq). Nearly all of these samples are derived from blood thus eliminating the problem of cell line artifacts. Using ensemble-based variant calls from read mapping and long-read genome assembly, we built a highly sensitive variant call set spanning the spectrum of variant size and complexity. From these calls, we built a map of haplotypes and recombination amongst the 10 second and third generation family members (two parents and eight children). Integrating the high-resolution haplotype map with multiple variant callers across sequencing technologies, we have built a truth set for a ten-member pedigree. In total, our pedigree-validated truth set contains 5,023,261 SNVs, 1,009,108 indels and 20,572 structural variants totaling 16 Mb of genetic variation.

By using inheritance patterns to validate the accuracy of the variant calls, this benchmarking database combines the strengths of the different technologies increasing the number of small variants by 14% and 6% in NA12878 compared to the Genome in a Bottle and Platinum Genomes, respectively. Compared to GIAB's NA12878, we have expanded the high-quality regions from 82.9% of the genome to 91.4%. Using technical replicates, we evaluate the accuracy of different sequencing technologies and variant callers against this comprehensive dataset.

The full sequencing data and validated variants identified in this study will be publicly available to serve as a valuable community resource, as the largest multi-generational pedigree sequenced with long-read technologies.

Keywords: DNA, Visual Impairment, Compiler, HealthCare

cDNA synthesis from ultra-low inputs of highly degraded RNA samples with novel RTase: Unlocking the potential for both low input and low-quality samples

Noreen Karim[1], Ekaterina Star[1], Berlin DeGroen[1], Michael Salgado[1], Sarah Beaudoin[2], Christopher Vakulskas[2], Benli Chai[3], Katica Ilic[1], Peilin Chen[1]

1 Integrated DNA Technologies, Inc., Redwood City, CA, 94065
2 Integrated DNA Technologies, Inc., Coralville, IA, 52241
3 Integrated DNA Technologies, Inc., Ann Arbor, MI, 48103

cDNA synthesis is one of the first and most critical steps for many molecular biology and genomic workflows including gene expression studies, functional genomics, comparative genomics, and single cell analyses among others. Synthesizing cDNA from low-quality RNA and/or formalin-fixed paraffin-embedded (FFPE) samples is often challenging due to the chemical modifications undergone during the preservation process reflected as very low yields and highly degraded extracted RNA.

We have developed modules for 1st and 2nd strand cDNA synthesis capable of preparing cDNA successfully used for NGS library preparation not only from very low RNA inputs (50 pg) but from highly degraded samples (RIN ⩾0.5) as well. The 1st strand cDNA synthesis module employs a novel reverse transcriptase and buffers for both fragmented and unfragmented RT workflows and yields at least 15-20X more cDNA compared to other commercially available kits. Moreover, our new 2nd strand cDNA synthesis following the proprietary 1st strand module consistently showed at least 2.5-fold higher yields compared to other competitors tested at RNA inputs ranging from 10 pg to 100 ng. The cDNA synthesized from variable quality samples including FFPE RNA, tissue RNA, and high-quality Reference/Standard RNA performed well when subjected to downstream applications such as library preparation using IDT's xGen cfDNA and FFPE DNA V2 MC library prep kit and hybridization capture using IDT's xGen Hybridization and Wash kit V2 following library prep. The sequencing results from the FFPE RNA samples using a 56-fusion gene hyb-capture panel showed the number of gene detected comparable to that from both tissue and reference RNAs, and detection of all fusion events anticipated in the samples. Moreover, the 1st stand cDNA synthesis module was successfully used for rRNA-depleted RNA samples to generate RNASeq libraries thereby showing compatibility with a variety of upstream and/or downstream processes.

(continue to page 169)

(continued from page 168)

Data suggests that our newly developed cDNA synthesis modules have high RT efficiency and can synthesize cDNA from highly degraded RNA (RIN >0.5) with inputs as low as 50 pg. Moreover, the modules can be effectively used for diverse molecular biology workflows, thereby providing an invaluable tool for biomedical and genetic research.

For research use only. Not for use in diagnostic procedures. Unless otherwise agreed to in writing, IDT does not intend for 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. RUO23-2437_001

Cell boundary segmentation for the Xenium Analyzer

Technology & Application Development

POSTER 537

Xiaoyan Qian[1], Dongyao Li[1], Daniel Gyllborg[1], Hsin Wen Chang[1], Hayato Ikoma[1], Paul Ryvkin[1], Neil Weisenfeld[1], Veronica Gonzalez Munoz[1]

10x Genomics, Pleasanton, CA.

10x Genomics Xenium Analyzer is a powerful platform for in situ analysis that enables users to query expression patterns of hundreds of genes, identify cell types and their microenvironment within a tissue, and how they arrange spatially to form tissue architecture. Cell segmentation is necessary to assign detected transcripts to their cells of origin. Xenium Analyzer has offered state-of-the-art nucleus segmentation, followed by isotropic expansion to approximate cell boundaries and assign transcripts. Here, we describe how we have identified stains that are best suited to cell segmentation across mouse and human tissues. We developed deep learning models to segment cells, with the goal of assigning transcripts more accurately for better gene expression quantification at single cell resolution.

We combined multiple stains into a single cocktail. We applied the segmentation cocktail and algorithm to a wide range of mouse and human formalin fixed & paraffin embedded (FFPE) and fresh-frozen (FF) tissue sections, including breast, brain, colon, liver, pancreas, kidney, and skin. In the liver, we recapitulated polyploidy and zonation. In the healthy lung, compared to isotropic expansion, alveolar cells matched better known cell morphology and shapes. In the brain, we could reduce the amount of misassigned transcripts coming from neuropils and more accurately identify rare cells in cortical layers.

The general-purpose segmentation solution that we demonstrate here is applicable to a wide range of tissues, and is additive to the updated Xenium workflow, providing a highly streamlined solution.

Clinical validation of mtDNA variant calling from WGS using DRAGEN

Areesha Salman[1], Diana M. Toledo[1], Marina DiStefano1, Andrea Oza[1], Katherine Lafferty[1], Edyta Malolepsza[1], Fontina Kelley[1], Teni Dowdell[1], Justin Abreu[1], Niall Lennon[1], Heidi L. Rehm[1,2]

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

The use of whole genome sequencing (WGS) to interrogate the mitochondrial genome (mtDNA) has become increasingly common, due to the ability to detect low-level heteroplasmic variation and the convenience of conducting analysis alongside analysis of nuclear DNA (Davis et al., 2022). Here, we describe the clinical validation of mtDNA variant calling and heteroplasmy determination using WGS data from the Illumina NovaSeq short read sequencing platform (PCR-free) with DRAGEN (v3.10.4) post-analytical processing, for use in the clinical WGS product at Broad Clinical Laboratories (BCL).

The ongoing validation study uses ~20 clinical and reference samples with known pathogenic mtDNA variants. Known variants include small variants (SNVs and InDels) and CNVs. Heteroplasmy levels for known variants were determined using an external CLIA-certified, CAP-accredited reference laboratory. Each sample was sequenced using existing clinical pipelines at BCL and met clinical sequencing quality specifications. Data was aligned to hg38 reference genome, mitochondrial variants were called against the revised Cambridge Reference Sequence (rCRS), and the resulting VCF was reviewed.

Validation metrics include sensitivity, specificity, repeatability and precision for mtDNA variant calling, and sensitivity, repeatability, and precision for determination of heteroplasmy levels of known mtDNA variants. The validation also includes the sensitivity and specificity of different mtDNA variant filtration protocols, implemented in Fabric Genomics. These protocols are used to identify variants in mtDNA for manual review and classification and are primarily based on heteroplasmy and gnomAD allele frequency.

This clinical validation will provide insight into the advantages and limitations of combining WGS analysis with mtDNA variant calling and heteroplasmy determination. If successful, we expect that this can help further streamline the process of providing a diagnosis to patients and reduce the need for multiple, incremental tests.

Comparative analysis of long-read approaches for single-cell transcriptomics

Iain Macaulay[1], Anita Scoones[1], Charlotte Utting[1], Lydia Pouncey[1], Ashleigh Lister[1], Naomi Irish[1], Karim Gharbi[1], Yuxuan Lan[1], David Swarbreck[1], Neelam Mehta[2], Adam Cribbs[2], Wilfried Haerty[1], David Wright[1]

1 Earlham Institute, Norwich, UK
2 Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford

Long-read sequencing can be used to incorporate isoform-level expression into single-cell transcriptomic studies, and although insightful, these approaches have typically been costly and yielded limited data for each individual cell. Recent advances in library preparation approaches and sequencing throughput have brought long-read single-cell studies closer to the mainstream.

In this study, we have performed a comprehensive, comparative analysis of several current and emerging approaches for single-cell long-read sequencing. Using peripheral blood mononuclear cells (10X Genomics), we have performed parallel analyses using the same cDNA to sequence via conventional (Illumina) sequencing, concatemerisation (MAS-seq) on Pacific Biosciences (Sequel IIe and Revio) and direct sequencing Oxford Nanopore platforms. Additionally, we have included CRISPR based depletion of highly expressed transcripts (Jumpcode CRISPRclean) as part of this experimental framework.

These parallel analyses from single-source cDNA libraries enables a direct comparison of each approach, in terms of standard metrics (i.e. cells captured, reads, genes and isoforms per cell) but also concordance in clustering and cell type identification. We also explore the ability of each approach to detect known and novel transcripts per gene and compare sequence-level variation between approaches including the detection of single-nucleotide variants and allelic expression. We also present an overview of the practical and cost implications of each approach.

In addition to benchmarking the performance of the methods in these areas, we focus specifically on immune cell receptors, T-cell receptors (TCR) and B-cell receptors (BCR) – we demonstrate that long-read sequencing of single-cell libraries can uniquely enable detection of all classes recombined TCR/BCR molecules without additional library preparation, including both alpha-beta and gamma-delta T-cells/TCRs.

Comparison between low-cost library preparation kits for low pass sequencing

Caitlin Stewart[1], Matthew Gibson[1], Gillian Belbin[1], Jahan-Yar Parsa[1], Jeremy Li[1], Geoff Benton[1], Tomaz Berisa[1], Joe Pickrell[1]

1 Gencove, New York, NY, USA.

The use of whole genome sequencing (WGS) to interrogate the mitochondrial genome (mtDNA) has become increasingly common, due to the ability to detect low-level heteroplasmic variation and the convenience of conducting analysis alongside analysis of nuclear DNA (Davis et al., 2022). Here, we describe the clinical validation of mtDNA variant calling and heteroplasmy determination using WGS data from the Illumina NovaSeq short read sequencing platform (PCR-free) with DRAGEN (v3.10.4) post-analytical processing, for use in the clinical WGS product at Broad Clinical Laboratories (BCL).

The ongoing validation study uses ~20 clinical and reference samples with known pathogenic mtDNA variants. Known variants include small variants (SNVs and InDels) and CNVs. Heteroplasmy levels for known variants were determined using an external CLIA-certified, CAP-accredited reference laboratory. Each sample was sequenced using existing clinical pipelines at BCL and met clinical sequencing quality specifications. Data was aligned to hg38 reference genome, mitochondrial variants were called against the revised Cambridge Reference Sequence (rCRS), and the resulting VCF was reviewed.

Validation metrics include sensitivity, specificity, repeatability and precision for mtDNA variant calling, and sensitivity, repeatability, and precision for determination of heteroplasmy levels of known mtDNA variants. The validation also includes the sensitivity and specificity of different mtDNA variant filtration protocols, implemented in Fabric Genomics. These protocols are used to identify variants in mtDNA for manual review and classification and are primarily based on heteroplasmy and gnomAD allele frequency.

This clinical validation will provide insight into the advantages and limitations of combining WGS analysis with mtDNA variant calling and heteroplasmy determination. If successful, we expect that this can help further streamline the process of providing a diagnosis to patients and reduce the need for multiple, incremental tests.

Comparison of enzymatic and automated acoustic DNA fragmentation for whole exome library prep from low input and degraded FFPE samples

Technology & Application Development

POSTER 541

Rachel Rock[1], Rebecca Beatty[1], Melanie Robinson[1], Dineen Wildman[1], Michael Sykes[1], Boris Umylny1, Thomas Halsey[1], Nripesh Prasad[1].

Discovery Life Sciences, Huntsville, Alabama 35756, USA.

DNA Fragmentation is the most critical step in each DNA library prep method for all short read-based protocols. Over the years several methods for DNA fragmentations have been made available, these methods are broadly classified into either Mechanical or Enzymatic categories. One of the most prevalent mechanical shearing techniques is acoustic shearing done by Covaris, where sound waves are used to induce cavitation in the sample which mechanically fragments the DNA. Enzymatic fragmentation cleaves the DNA at different points depending on the chosen enzymes. The benefits and drawbacks of each method must be considered for samples to optimize workflow. Here we report a comparison of enzymatic shearing using NEB Ultra II FS enzyme (NEB) and a fully automated Covaris acoustic mechanical shearing on NA12878 and degraded FFPE samples in triplicate at 200ng, 100ng, 50ng, and 10ng. The library prep on these sheared samples was performed using the components of the NEB Ultra II FS kit for Whole Exome Sequencing. DNA libraries from the two shearing methods were captured using xGen™ Exome Hyb Pane lv2 (IDT). The libraries were sequenced on NovaSeq6000 as 100bp PE and 150-300X overall coverage. We evaluated the libraries based on the average library insert size, Q30 base %, Duplicate %, and Discordant %. We observed larger insert size libraries of the NA12878 sample with Covaris shearing (~241bp) as compared to enzymatic shared DNA (~218bp). Q30 base % was higher for Covaris sheared libraries (~94.5%) as compared to enzymatic sheared (~93.5%) for both NA12878 and FFPE libraries at all input levels. We observed higher duplication rates with both Covaris and enzymatic sheared libraries at 10ng input. Observed discordance rates were much higher in enzymatically sheared DNA libraries (7-13%) than Covaris sheared (average <1.6%) for NA12878 samples at all DNA inputs. This points to higher error calls and thus technical artifacts affecting robust genotyping calls with enzymatic shearing. Enzymatic fragmentation has the advantage of low upfront costs and streamlining of the workflow. However, taken together, our testing shows that mechanical shearing outperforms enzymatic shearing in both degraded FFPE and NA12878 samples, especially with limited input DNA material.

Comparison of three different DNA library prep kits for whole exome sequencing optimization from low input and degraded FFPE samples.

Nripesh Prasad[1], Rachel Rock[1], Rebecca Beatty[1], Melanie Robinson[1], Dineen Wildman[1], Sarah Landrum[1], Michael Sykes[1], Boris Umylny[1], Thomas Halsey[1].

Discovery Life Science, Huntsville, Alabama 35756, USA.

Formalin-fixed paraffin-embedded (FFPE) methods have been utilized to preserve clinical samples for long terms and at low cost, thereby providing an opportunity for expanding discovery and diagnosis by deploying next-generation sequencing methods. However, the FFPE preservation process leads to the degradation of the nucleic acids. This creates a need for a robust library protocol to utilize degraded FFPE samples at sub-50ng input amounts for whole exome sequencing. Here we report a comparison of three library preparation kits for low-input, degraded FFPE samples: Kapa HyperPrep (Roche), xGen™ ssDNA Low-input DNA Prep (Swift/IDT), and Ultra™ II FS DNA Library Prep (NEB). For each library prep kit, FFPE samples were tested in triplicate at 200ng, 100ng, 50ng, and 10ng, respectively. Pre-capture libraries were evaluated to determine the total DNA library available, with a goal of at least 600 ng to 100 ng for hybridization. DNA libraries from all three prep methods were captured using xGen™ Exome Hyb Panel v2 (IDT). The libraries were sequenced on NovaSeq6000 as 100bp PE and 150-300X overall coverage. The performance of each library preparation kit was evaluated on the number of libraries eligible for capture, percent duplication rate, median insert size, and coverage. NEB had the most eligible libraries reaching 600ng and only one library falling below 100ng. Roche and Swift/IDT performed similarly to each other with 22 and 24 libraries falling below 100ng. The duplication rate for NEB ranged from 25-50% at 50ng input, the Swift prep reached an 80-90% duplication rate at similar inputs. Similar trends are found with median insert size, with NEB generally producing >200bp with Roche and Swift/IDT producing shorter insert libraries <200bp. NEB also outperformed with inputs as low as 50ng for highly degraded FFPE samples and 10ng for moderate quality FFPE samples achieving ~150x coverage. In comparison, Swift/IDT failed to reliably meet 150x coverage and Roche met coverage only with moderately degraded samples. Taken together, our testing shows that Ultra™ II FS DNA Library Prep by NEB outperforms other kits tested with highly degraded FFPE samples, especially those with limited input DNA material.

Comprehensive characterization of transcript isoform in the human and mice retina with single cell long-read RNA sequencing

Rui Chen[1, 2], Meng Wang[1], Yumei Li[1], Soo Hwan Oh[1], Xuesen Cheng[1], Jun Wang[1]

1 Baylor College of Medicine, Department of Molecular and Human Genetics, Houston, TX, USA
2 Baylor College of Medicine, Human Genome Sequencing Center, Houston, TX, USA

As a complex neuronal tissue, the retina comprises over 100 cell types, which can be distinguished based on morphology, function, location, and transcriptomic profile. Currently, the extent of mRNA alternative splicing in the retina has yet to be systematically characterized at the individual cell type level. To address this gap, we utilized single-cell RNA sequencing, combining both short-read and long-read high-throughput techniques. The transcriptomes of around 30,000 mouse retina cells have been profiled, totaling 1.54 billion Illumina short reads and 1.40 billion Oxford Nanopore long reads. Strikingly, 44,325 transcript isoforms have been identified, with 55% are novel and 17% showed cell class-specific expression. Interestingly, although many isoforms were expressed across various cell types, their distribution often varied among types. Therefore, this research significantly expands the transcriptome isoform catalog and sets the stage for further exploration of alternative splicing in retinal biology and its related diseases. Concurrently, we profiled around 100,000 human retina cells using a similar approach. A comparison between human and mouse retina transcriptomes is currently underway, with updates to follow.

Deep, targeted simultaneous sequencing of genetics and epigenetics in DNA

Technology & Application Development

POSTER 544

Jack M Monahan, Aurel Negrea, Lisabet Andreasen, Paula Golder, David J Morley, Rosie Telford Spencer, Minna Taipale, Jens Fullgrabe, Walraj S. Gosal, Páidí Creed

biomodal Ltd, Cambridge, United Kingdom

DNA comprises molecular information stored in genetic and epigenetic bases, both of which are vital to our understanding of biology. The interplay between genetics and the DNA epigenome orchestrates complex biological phenomena as diverse as cell fate, ageing, the response to environmental stimuli, and disease pathogenesis. Methods widely used in the detection of methylated cytosines fail to discriminate between thymines and unmodified cytosines. Consequently, they fail to capture common C-to-T transitions and provide incomplete genetic information. The duet multiomics solution is a new technology which sequences at single base resolution the complete genetic sequence integrated with modified cytosine.

Whole-genome sequencing with duet multiomics solution generates high quality genetic and epigenetic information, even from low DNA input. This enables the identification of genetic variants and quantification of methylated cytosine levels in a single experiment. The phased nature of the technology, whereby genetic and epigenetic information is available jointly at a read-level, enables the study of genetic and epigenetic co-variation.

Targeted sequencing enables high-resolution measurement of variation in a subset of the genome, where sequencing to a comparable depth would not be feasible for whole-genome approaches. Here we show the compatibility of duet sequencing with hybrid capture using commercially acquired probe sets and demonstrate that the enrichment obtained is comparable to that obtained with other epigenetic sequencing methods while maintaining the high accuracy of duet multiomics solution. A single hybrid capture with duet multiomics is sufficient to obtain accurate genetics and epigenetics for regions of interest and avoids the otherwise complex data integration that is necessary with alternative epigenetic NGS techniques. Targeted sequencing with duet multiomics solution can be used to disentangle and dissect the genetic and epigenetic variation in mosaic genomes and heterogeneous populations of cells by phasing genetic variants and cytosine methylation.

Deeper characterization of immune repertoire diversity observed with longer read sequencing and SBS accuracy improvements

Technology & Application Development

POSTER 545

Carolyn G. Conant[1], Alix Kwasniewska[2], Ilaria Grizzi[2], Tim Merkel[2], Camilla Columbo[2], Cande Rogert[1]

1 Illumina Inc., Core R&D, San Diego, California, USA
2 Illumina Inc., Core R&D, Cambridge, Cambridgeshire, UK

Immune repertoire sequencing requires longer reads to capture full-length clonotypes via enhanced pair-read overlap which is not achieved with commonly used 300-cycle next-generation sequencing. Additionally, higher data quality at the ends of the reads confers the ability to discover larger numbers of unique clonotypes due to better mapping fidelity. We demonstrate improved clonotype detection using 600-cycles, higher quality reads enabled by XLEAP-SBS™ chemistry on a NextSeq™ 1000/2000 platform. We observe a marked increase in throughput, up to 33% higher, and a major reduction in sequencing error rates, up to 50% lower. Detection of clonotypes is significantly improved with this configuration. Scaling complex IR-Seq experiments with higher quality deep reads results in a more comprehensive view of the diversity and clonality of B-cell IgG and T-cell receptors.

Defining the transcriptomic identity of PIK3CA-mutated cells in a human capillary malformation using single cell long-read sequencing

Katherine E. Miller[1,2], Jesse J. Westfall[1], Jason B. Navarro[1], Maria Elena Hernandez Gonzalez[1], Elaine R. Mardis[1,2], Catherine E. Cottrell[1,2,3], Anthony R. Miller[1], & Michelle A. Wedemeyer[1,2,4]

1 Abigail Wexner Research Institute at Nationwide Children's Hospital, Institute for Genomic Medicine, Columbus, OH, USA
2 The Ohio State University College of Medicine, Department of Pediatrics, Columbus, OH, USA
3 The Ohio State University College of Medicine, Department of Pathology, Columbus, OH, USA
4 Nationwide Children's Hospital, Department of Neurosurgery, Columbus, OH, USA

Somatic mosaicism is a well-known contributor to vascular malformations, and research in this domain has been steadily growing. In this study, we employed single-cell RNA sequencing, encompassing both short- and long-read technologies, to link the genotype of a known mosaic PIK3CA mutation with the transcriptional profile in more than 10,000 cells isolated from a human capillary malformation. Our patient exhibited symptoms indicative of megalencephaly-capillary malformation-polymicrogyria syndrome, prompting a skin biopsy of the affected area. Exome sequencing of blood revealed no pathogenic mutations, but sequencing of cultured biopsy tissue unveiled a PIK3CA mutation (c.3139C>T:p.His1047Tyr; variant allele frequency 11.9%), classified as a Tier I (Level A) variant. (PMID: 27993330).

While droplet-based single cell RNA sequencing allows for gene expression analysis in thousands of cells, it poses limitations in simultaneous genotyping given that sequence information is provided only for short fragments at the transcript ends. To profile both the transcriptome and PIK3CA mutational status of the single cells, we employed both short-read (10x Genomics Single Cell 3' Gene Expression kit) and long-read (PacBio MAS-Seq 10x Single Cell 3' kit) technologies. After quality control, 12,378 cells were included in downstream analysis. Long-read sequencing provided efficient genotyping for 1,894 of these cells, and 18.7% of those harbored the PIK3CA c.3139C>T mutation. After mapping to a reference skin set (PMID: 36732947), most cells were categorized as fibroblasts with the remaining 10% identified as undifferentiated keratinocytes or pericytes. The first principal component was primarily influenced by the PIK3CA mutation status, while the second principal component was primarily influenced by the cell cycle status.

(continue to page 180)

(continued from page 179)

The most differentially expressed genes in PIK3CA-mutant cells were associated with inflammation (CXCL14), cytoskeleton (TUBB2B), chromatin remodeling (BC11B), cell-cell interactions (NCAM1), and transcriptional regulation (PAX3, NFIB). Elevated expression of PAX3 and NCAM1 in the PIK3CA-mutated cells suggests they exhibit characteristics of immature neural crest cells. Our results demonstrate successful utilization of single cell technologies to enable high throughput linking of genotypic and transcriptomic data.

Detection and subtyping of multiple Salmonella serovars from complex samples using precision metagenomics

Christopher Grim, Padmini Ramachandran, Kranti Konganti, Amanda Windsor, Elizabeth Reed

Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration

Foodborne pathogens, such as Salmonella, are often low abundant species in a high biomass microbiome. Routine surveillance for Salmonella is often reliant on culture-based amplification and results in detecting the most abundant serovar. Here, we introduce a precision metagenomics workflow for Salmonella subtyping from shotgun metagenomic or quasi-metagenomic (enriched metagenomes) datasets, that combines bettercallsal, a novel bioinformatics pipeline, and SalSUH (Salmonella Subtyping Using Hybridization), a novel enrichment panel targeting 32% of the pangenome including specific diagnostic markers for 30 clinically relevant serovars.

bettercallsal was evaluated using in-silico, retrospective outbreak, and benchmark samples. In-silico datasets, comprising 29 unique Salmonella, 46 non-Salmonella bacterial and 10 viral genomes, were generated with read depths from 0.5 million to 5 million read pairs. For outbreak samples, previously sequenced quasi-metagenomic datasets, 24hr pre-enrichments and selective enrichments for Salmonella, from papayas that yielded multiple Salmonella serovars were used. Benchmark datasets of up to six Salmonella serovars were created by combining equal volumes of each serovar into a high background of enriched soil microbiome at three levels: 10% (high), 1% (low), or 0.1% (ultra-low). Additionally, single serovar dilution series in the same background were constructed to determine the limit of detection. Libraries were sequenced with and without hybridization to the bait panel and analyzed with bettercallsal.

The in-silico dataset analyzed with bettercallsal revealed that accuracy, recall, and specificity improved with increased read depth, and correctly identified up to 27 out of 29 serovars. For retrospective datasets, bettercallsal identified multiple serovars in concordance with the original outbreak findings and the genome hits assigned by bettercallsal match the isolate sequences from the papaya outbreak as evident by SNP clustering, in NCBI Pathogen Detection. SeqSero2 typically yielded partial antigenic profiles or single serovars. For ultra-low benchmark samples, SalSUH coupled with bettercallsal identified 80-100% of multiple serovars in a single sample compared to only one serovar detected from the same libraries sequenced without hybridization. SeqSero2 and kmer analyses were not able to correctly identify mixed serovars with or without hybridization. Further, the combined precision metagenomics approach demonstrated a limit of detection of less than 100 genome copies in this high biomass background.

Determination of cell viability using high-dimensional morphology analysis

Anastasia Mavropoulos, Emilie Schneider, Jordan Nieto, Ryan Carelli, Kiran Saini, Amy Wong-Thai, Stephane Boutet, Senzeyu Zhang, Vivian Lu, Mahyar Salek, Maddison Masaeli

Morphological changes associated with alterations in cell state, health, and fitness have been reported in drug response, cell differentiation, and disease progression studies. However, the ability to profile sample health at scale can be costly, time consuming, and labor intensive. Here, we utilized label-free morphological profiling of cells on the REM-I platform to determine cell viability. Deepcell's REM-I is a research platform for phenotyping and sorting cells of morphologically similar profiles based on real-time deep learning interpretation of high-content morphology data without biomarker labels.

Healthy and viable cell line samples were subjected to various chemical and physical methods to induce cell death. The varied resulting cell health states were imaged on the REM-I platform as single cell suspensions. The captured brightfield images of single cells were characterized using the Deepcell Human Foundation Model (HFM) to infer morphological features associated with respective cell health states. The HFM is a hybrid architecture that combines self-supervised learning and morphometrics (computer vision) to extract 115 embedding vectors representing cell morphology from the high-resolution REM-I images. High-dimensional morphological profiling showed that untreated, apoptotic, and necrotic cells exhibited distinct morphotypes, with respective cell death pathways localizing to distinct embedding space. This data suggests that label-free profiling using deep learning and morphometrics can be used to determine cell viability and state of health.

Next, we applied respective embeddings extracted from images of viable and non-viable cells to train a random forest classifier to infer features linked to cell viability. A total of 1.5 million images of viable and non-viable cells acquired in various cell lines compose the training set. Model performance evaluation of the 'Cell Viability Classifier' on the hold-out dataset showed over 97% specificity and 96% sensitivity. Combined with the REM-I platform, the 'Cell Viability Classifier' can be used to determine the viability of cells in a sample without stains and dyes. Importantly, cells of interest can be sorted and collected for downstream assays requiring high-quality, viable cells, such as single cell RNA-sequencing, cell culture expansion, and functional assays.

Developing Multiple Novel Methods for the Generation of Single-Nucleus RNA-Seq Data from Frozen PAXgene Blood

Ojasvi Chaudhary[1], Grant Duclos[1], Peter Gathungu[1], Manisha Rao[1], Rogelio Aguilar[1], Liat Amir-Zilberstein[1], Varsha Shankarappa[2], Chris Rands[2], Xi Chen[3], Rebecca Halpin[3], Joseph Boland[3], Maurizio Scaltriti[3], Brian Dougherty[1], Asaf Rotem[1]

1 AstraZeneca, Oncology R&D, Translational Medicine, Waltham, MA, USA
2 AstraZeneca, Oncology R&D, Oncology Data Science, Cambridge, UK
3 AstraZeneca, Oncology R&D, Translational Medicine, Gaithersburg, MD, USA

Blood is commonly collected for molecular testing directly into "PAXgene Blood RNA" (Qiagen) solution, which permeabilizes cells and stabilizes nucleic acids for long-term, archival storage. RNA extracted from PAXgene blood has been used for bulk transcriptomic profiling, but it has not been shown that these samples can be profiled at single-cell resolution. Therefore, we have developed two methods for single-nucleus (sn) RNA-Seq of frozen PAXgene blood. The first method, "Cell Lysis" (CL), involves detergent-based lysis of pelleted blood for nuclei extraction, whereas the second method, "Mechanical Separation" (MS), involves filter-based separation of white blood cell nuclei from the remainder of the blood sample. In this study, fresh blood from three healthy donors was stored and frozen in PAXgene solution, then nuclei were isolated from replicate specimens using both the CL and MS methods (n=12 samples total). 10x Genomics' Chromium platform was then used to generate 3'RNA-Seq data from roughly 5,000 nuclei per sample. The CL method produced significantly higher ($p<0.05$) nuclei yield, relative to MS (10:1 CL:MS). However, significantly higher ($p<0.05$) levels of estimated ambient RNA contamination was detected in nuclei when using the CL method, relative to MS (4:1 CL:MS). Importantly, a CRISPR-based strategy was implemented to deplete specific, high-abundance, ambient transcripts from sequencing libraries, which significantly reduced ($p<0.05$) estimated contamination to negligible levels in nuclei in downstream snRNA-Seq data. Furthermore, unsupervised clustering was used to define groups of cells, which were assigned cell types based on marker gene expression. In addition to expected mononuclear cell populations, such as T cells, B cells, and monocytes, distinct clusters of granulocytes (neutrophils, basophils, eosinophils) were detected in all samples. Interestingly, it was found that the MS method yielded a significantly higher ($p<0.05$), proportion of granulocytes than the CL method, but mononuclear cell type proportions were consistent between the two methods. Findings from this study have led to the development of two novel strategies for snRNA-Seq data generation from blood frozen in PAXgene solution. These results represent an important step forward in expanding our capability to apply single-cell genomic technologies to a broader array of archived biospecimens from clinical studies.

Development and scaling of a sensitive NGS-based Assay for clinical diagnosis of acute meningitis and encephalitis

Peter Trefry[1], Tim Desmet[1], Niall Lennon[1], Justin Abreu[1], Michael Dasilva[1], Wendy Brodeur[1], Evan McDaid[1], Raya DePina[1], James Harrold1, Cori Milbury[2], Emily Crawford[2], Anna Uehara[2], Fay Wang[2], Steve Miller[2], Brian O'Donovan[2], Nick Gimbrone[2] , Andrew Carretta[2]

1 Broad Clinical Labs, Cambridge, MA, USA.
2 Delve Bio Inc. Boston, MA, USA

Meningitis and encephalitis are two forms of acute neurological illnesses that can lead to long-term disabilities or even death. Currently, these infections are diagnosed through a differential diagnostic approach which leverages the patient's history, clinical representation, imaging, and serial laboratory tests. This traditional approach can be time-consuming, costly, and only identifies the causative pathogen in approximately 50% of acute cases. Additionally, many of these tests often come back negative, resulting in prolonged time between admission to diagnosis and associated treatment course. Leveraging next-generation sequencing provides an opportunity to create a single, highly-sensitive pathogen-agnostic assay for routine clinical care, but attempts to do so have been restrained by difficulties scaling the workflow, high costs, long turnaround time, and challenges with clinical interpretation.

DelveBio and the Broad Clinical Labs have collaborated to develop a scalable and sensitive metagenomic NGS-based workflow that provides a diagnosis in a timeframe equivalent to conventional send-out testing. Cerebral spinal fluid is processed through a fully automated workflow, generating dual-indexed Illumina-compatible sequencing libraries derived from both DNA and RNA. Single-end sequencing is performed on the resulting libraries using the Illumina NovaSeq 6000 instrument on an SP flowcell. Data is then processed through Delve's analysis pipeline, which aligns non-human reads to a curated version of the Genbank database of all known and potential pathogens.

Here Broad Clinical Labs and DelveBio present a sensitive, scalable, and fully automated diagnostic workflow which enables rapid pathogen-agnostic screening for clinically relevant and causal pathogens of acute neurological illnesses. Leveraging NGS has expanded the number of pathogens screened for in a single test, reducing the number of tests required to correctly diagnose patients, while maintaining a turnaround time similar to or better than current send-out testing. By connecting physicians with this technology, patient outcomes are improved via shortened hospital stays and reduced time between admission and treatment, both of which reduce overall costs.

Development of a high-throughput NGS library preparation workflow with normalization adapters and in-line barcode technology

Owen Smith, Kristin Butcher, Sean Tighe, Su Maw, Tong Liu, Cibelle Nassif, Danny Antaki, Esteban Toro, Elian Lee, Ramsey Zeitoun, Siyuan Chen

Twist Bioscience, Research and Development, South San Francisco, California, USA

Advancements in next-generation sequencing (NGS) continue to decrease sequencing costs by increasing data generated per run. This reduction in sequencing cost enables population-level genomic experiments to help study and diagnose genetic disorders. Despite sequencer advancements, efficient NGS library preparation is hindered by the cost and effort needed to process individual samples. Libraries need to be quantified and pooled to equitably distribute sequencing reads, which becomes prohibitively tedious, costly, and time-consuming as sample numbers increase. We present a streamlined NGS library preparation technology that eliminates the need for sample-by-sample handling normalization, and pools samples early in the process, resulting in great simplification in workflow and up to a 100x increase in post-ligation sample processing throughput without compromising data quality.

This high-throughput workflow is built upon enzymatic fragmentation of samples that are converted to sequenceable libraries using novel normalization adapters with in-line barcodes. Library normalization occurs from adapters during the ligation step, achieving library conversion independent from DNA input mass (20-200 ng, CV < 20%) on-par with qPCR-based pooling. Up to 96 in-line barcodes are included on the adapters which uniformly ligate to samples for simple demultiplexing. The design of the in-line barcodes allows for flexible pooling where up to 96 samples can be pooled in a modular manner. With normalization adapters and in-line barcodes, pooling is performed immediately after adapter ligation, significantly reducing the number of clean-up and PCR reactions. This format can also support up to tens of thousands of libraries in a single sequencing run using dual unique index sequencing primers post-ligation. Finally, multiplexed PCR amplification artifacts can be reduced using this workflow. Library conversion is consistent and well-suited for both low-pass WGS and targeted sequencing for variant calling (>20x coverage).

This new library preparation process applies a novel technology to reduce hands-on time and increase efficiency. A process that once treated each sample individually can now be pooled while ensuring equal conversion independent of DNA mass input. This workflow significantly advances the experimental process of library preparation to complement throughput advancements made in sequencing instrumentation.

Digital Fluidics enables broad-based access to push-button NGS sample prep

Technology & Application Development

POSTER 552

Abdul Mohammed, Shikhar Gupta, Jaiden Cruger, Alice Wong, Carlos Ruiz-Vargas, Bella Barbera, Pierre Sphabmixay, Craig Broady, James Ferguson, Udayan Umapathi

Volta Labs, Inc., Boston, MA

Decades and billions of dollars have been dedicated to advancing sequencing technologies and genomic data analysis. However, the sample preparation process remains largely unchanged. When samples are processed manually, variable sample quality, laborious protocols, and high cost remain significant obstacles to the scalability and consistency of NGS sample prep. Automated liquid handlers enable higher throughput, but barriers to implementation persist. These include high capital expenditure, the need for specialized infrastructure and/or skills, lengthy method development and ongoing cost of maintenance.

Volta is introducing a reimagined Digital Fluidics platform that offers precise and programmable control, magnetic manipulation, contact-free mixing, and precise temperature control for liquids and paramagnetic beads used in NGS sample preparation. Our system accommodates a wide range of volumes, spanning two logarithmic units. Sample and reagent manipulation is largely plastic-free, minimizing material loss and preserving sample integrity. Tight control over enzymatic reactions and unique capabilities enable miniaturization, thereby reducing cost and turnaround time while preserving sequencing quality.

The Volta platform is capable of processing a raw biological sample all the way through sequencing-ready molecules ready for whole genome, whole exome, or targeted sequencing on any commercially available sequencer. DNA extracted on the platform is of a very high quality, with a tight distribution around a mode size >150 kb. Libraries prepared in miniaturized format with the PacBio SMRTbell Prep Kit 3.0, the Illumina DNA PCR Free Prep Kit, and enzymatic fragmentation/ligation chemistry match or exceed manufacturer's quality and sequencing metrics. For hyb-capture workflows, the platform eliminates extreme operational complexity and variability in sequencing metrics such as on-target rate and uniformity.

(continue to page 187)

Our sample preparation platform provides access to cutting-edge sequencing methods that outperform highly skilled technicians and simultaneously lowers costs through miniaturization and freeing up labor time. It substantially reduces the need for operator training in complex NGS sample processing, ensures fast turnaround times by offering flexibility in sample batch size – driving the dissemination of genomics in emerging clinical use cases. We have enabled push-button execution of a wide range of NGS sample prep protocols in batch sizes compatible with both benchtop and production scale sequencers.

Digital fluidics simplifies NGS library preparation in a clinical setting

Technology & Application Development

POSTER 553

Udayan Umapathi[1], Lorena Mora-Blanco[2], Nicholas Wong[2], Jennifer Corona[2], Daniel Giguere[1], and William Lee[2]

1 Volta Labs, Inc., Boston, MA
2 Helix, Inc., San Diego, CA2Helix, Inc., San Diego, CA

Next-generation sequencing (NGS) has become an established technology for molecular diagnostics. Automated liquid handlers are widely used for the preparation of NGS libraries to achieve process control, but require a significant investment and skilled personnel to implement and maintain, and typically reduce flexibility to process small batch sizes. In this study, we investigated the suitability of the Volta Labs Digital Fluidics Platform—which employs electrical and magnetic fields—for NGS library preparation in a clinical setting.

The Illumina DNA PCR-free Prep protocol (which employs bead-linked transposomes to generate auto-normalized, sequencing-ready libraries) was automated on a Volta instrument in miniaturized (½ volume) format. Libraries were prepared from reference and human genomic DNA, sequenced to ⩾30X coverage on an Illumina NovaSeq X sequencer, and analyzed using Helix's internal data analysis pipeline. Analysis of library QC and sequencing data confirmed that miniaturized Volta libraries met the yield and quality requirements for Helix's standard WGS pipeline. Variant calling metrics for reference and clinical samples were equivalent to metrics for libraries prepared using Helix's current manual workflow.

A second focus of the study was to assess reproducibility between libraries, users, and runs. Pre-sequencing metrics (library fragment size distribution and yield) were assessed for libraries prepared on the Volta instrument across nine independent runs by three different operators. Representative libraries were sequenced at ⩾30X coverage on an Illumina NovaSeq X sequencer, and analyzed using Helix's internal data analysis pipeline. Output metrics were compared to data previously generated using Helix's clinical library preparation workflow.

Apart from yielding high-quality WGS libraries, the Volta platform was found to offer true walk-away library prep with easy setup and minimal sample handling, thereby reducing the risk of error. Since the miniaturized workflow requires lower input (⩾150 ng), it is also possible to repeat library preparation from limited clinical samples in case of error or failure at any point in the pipeline. These attributes of the Volta system provide significant benefits for NGS library preparation in a clinical setting.

Direct multi-omics for the masses: inking DNA methylation to chromatin targets via TEM-seq

James Anderson[1], Bryan J. Venters[1], Vishnu U. Sunitha Kumary[1], Allison Hickman[1], Anup Vaidya[1], Ryan Ezell[1], Jonathan M. Burg[1], Louise Williams[2], Chaithanya Ponnaluri[2], Hang Geong Chin[2], Pierre Esteve[2], Isaac Meek[2], Sriharsa Pradhan[2], Zu-Wen Sun[1], Martis W. Cowles[1], & Michael-Christopher Keogh[1]

1 EpiCypher Inc, Durham, NC 27709, USA
2 New England Biolabs, Ipswich, MA 01938, USA

DNA methylation (DNAme) is an epigenetic mark that includes the modification of cytosine resides (5mC) within CpG islands. In addition to well characterized roles regulating gene expression, imprinting and silencing parasitic DNA elements, the misregulation of DNAme is implicated in multiple diseases. Evidence is emerging that DNAme is not an independent epigenetic mark but rather closely linked to the post translational modification (PTM) of histone proteins. However, examining the direct relationships between 5mC and PTMs are hampered by correlated analyses of separate assays that cannot establish a direct mechanistic linkage. Furthermore, the traditional approach to measure 5mC relies upon harsh bisulfite chemical conversation of DNA, which introduces DNA breaks and systemic biases.

To address these limitations, we developed a Targeted Enzymatic Methylation-sequencing (TEM-seq) approach, an ultra-sensitive multi-omic genomic mapping technology that delivers high resolution DNAme profiles at epitope-defined chromatin features. Importantly this assay is capable of examining the direct molecular link between 5mC and histone PTMs and/or chromatin associated proteins (ChAPs). This multi-omic workflow integrates the CUT&RUN assay for genomic mapping together with NEB's enzymatic methyl-seq (EM-seq) for unbiased DNAme analysis. CUT&RUN is a highly sensitive, standards-driven assay that uses antibodies to locally tether protein A/G-micrococcal nuclease (pAG-MNase) to chromatin within intact cells or nuclei, followed by controlled activation of the MNase to cleave nearby DNA. EM-seq leverages enzymatic conversion of DNAme to generate single-base resolution, unbiased DNAme profiles from sub-nanograms amounts of CUT&RUN-enriched DNA.

(continue to page 190)

(continued from page 189)

To determine the capabilities and limitations of this assay, we tested multiple chromatin targets in various cell lines. We demonstrate that TEM-seq is highly reproducible, specific, and efficient (<10% off-target binding, >90% enzymatic conversion of DNAme, <0.5% conversion of unmethylated DNA). TEM-seq is also highly-sensitive, requiring only 5 million short-read sequences per assay (10-50x less than whole-genome 5mC-sequencing), demonstrating the disruptive potential of the assay to enable cost-effective approach for targeted DNAme analysis. Finally, we leveraged TEM-seq multi-omics capabilities in a clinically relevant model to gain mechanistic insights in a neurodegenerative disorder, Rett syndrome, that is the result of the mutations in the 5mC-binding domain of the MECP2 gene.

Discovery of single-cell mutations with multiplexed full-length transcriptomes and long read sequencing

Technology & Application Development

POSTER 555

Sharmili Roy[1], Susan M. Grimes[1], GiWon Shin[1,2], Billy T. Lau[1], Hanlee P. Ji[1,3]

1 Division of Oncology, Department of Medicine, Stanford University, School of Medicine, Stanford, CA, 94305, United States
2 Washington University School of Medicine in St. Louis, MO, 63130, United States
3 Department of Electrical Engineering, Stanford University, Stanford CA, 94305, United States

Single-cell sequencing identifies complex genomic characteristics of individual cancer cells even when the tumor is embedded among other cell types in the local tumor microenvironment. However, the identification of cancer point mutations among individual tumor cells is not available among most single cell genomic assays. The major challenges for single-cell multiplex technologies include (1) the need for a substantial amount of starting DNA samples for complex multiplexed enrichment of specific genomic targets, (2) many artifacts that hinder amplification efficiency often go undetected, and (3) requiring advanced analytical techniques and significant computational resources for accurate interpretation.

To address these challenges, we introduce a novel single-cell technology utilizing gene-specific primers (GSP) to iteratively target multiple genes from single-cell assays and identify mutations from long read sequencing without depletion of the source scRNA-seq library. We also utilize click chemistry to form a covalent bond between a single cell RNA-seq (scRNA-seq) library and a solid phase support such as magnetic beads or plastic PCR tubes. The solid phase support surface undergoes functionalization with tetrazine and trans-cyclooctene groups, facilitating a robust attachment for the DNA molecules from a scRNA-seq library. Conjugating the DNA molecules to the solid phase securely preserves the original sequencing library even during iterative polymerase-based extension reactions that copy the library. In contrast, other types of enrichment experiments would rapidly deplete the scRNA-seq source material.

(continue to page 192)

(continued from page 191)

We applied GSP-based multiplexed targeting of specific genes to increase the coverage of specific gene sets. This technology enables repeated sequencing while preserving the loss of original sequencing library. For sequencing specific gene targets, we maintain the intact full length single cell cDNA and use the long reads to cover the entirety of the mRNA sequence (in cDNA form). We applied a two-step nested amplification reaction and subject the library to long-read sequencing with an Oxford Nanopore sequencer. We successfully achieved at least 10-fold enrichment compared to the baseline expression of each gene target prior to GSP amplification. In summary, the GSP-based enrichment process, coupled with conjugation, offers highly multiplexed targeted amplification for identifying mutations and thus facilitating multi-level genomic analysis for cancer research.

DNA methylation-based CNS tumor classification with nanopore sequencing

Assaf grunwald[1,2], Galina Deinberg[1], Shai Izraeli[2] and **Yuval Ebenstein[1]**

1 Rina Zaizov Pediatric Hematology Oncology Division, Schneider Children's Medical Center of Israel, Petah Tikva, Israel.

2 School of Chemistry, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel

Central nervous system (CNS) tumors are the most common type of solid tumors in children, presenting a challenge due to the variety of tumor types and the difficulty in obtaining adequate tissue samples for diagnosis. Traditional diagnostic methods, such as pathology, may be slow and inconclusive, leading to delays in treatment planning. In a landmark study from 2018, DNA methylation profiling, generated by DNA arrays, was shown to be a reliable method for classifying CNS tumors. The study found that DNA methylation patterns are highly consistent within specific tumor types and that this technique could accurately classify tumors even when traditional diagnostic methods failed to do so. A random forest classifier (RFC) model was developed using a previously known cohort of CNS tumors as a training set and validated with an independent cohort, demonstrating high accuracy and reliability.

Oxford nanopore sequencing technology (ONT) enables rapid and accurate profiling of DNA methylation marks on native DNA molecules without the need for bisulfite conversion. In collaboration with the Schneider Medical Center, we are establishing an ONT-based solution for CNS classification utilizing the RFC model developed on DNA arrays. The assay was successfully validated on the first cohort of samples from patients. Our approach has the potential to revolutionize CNS tumor classification, allowing for faster and more accurate diagnosis, which in turn could lead to more effective and personalized treatment options, particularly in children.

e.cuppen@hartwigmedicalfoundation.nl

Edwin Cuppen

Whole genome sequencing (WGS) of tumor-normal paired samples is the most complete approach for comprehensive molecular diagnostics for solid tumors. Although clinically validated, wide-spread adaptation is still hampered by higher costs compared to panel-based approaches. In contrast to panel-based approaches, which involve relative expensive target enrichment reagents, the main cost component for WGS is in the sequencing itself. Therefore, sequencing technology innovations that reduce the costs of sequencing can have a very large impact on the overall test costs and hence broad clinical adoptation.

The UG100 sequencer from Ultima Genomics uses mostly natural sequencing-by-synthesis chemistry on an innovative open fluidics platform at a price tag of $1 dollar per Gigabase. With 90x tumor and 30x normal control genome coverage, reagent costs for WGS-based cancer diagnostics test is potentially reduced from a currently ˜1800-2000 dollar (Novaseq6000) to ˜400 dollar.

To assess the suitability of the UG100 platform for diagnostic purposes, we sequenced the well-studied COLO829 tumor-normal cell line pairs and performed in-depth recall and precision analyses as compared with Novaseq6000 results. Sequencing data was processed and genome-wide variation annotated with the Hartwig Medical Foundation open source data analysis pipeline for WGS-based cancer diagnostics. This pipeline was originally developed for Illumina paired-end sequencing data (2 x 150nt), but was slightly modified and optimized to handle UG100 data (single end ˜300nt).

We will report on detailed analyses for small variant (SNV, indel), copy number and structural variant detection, demonstrating high genome-wide accuracy and concordance with Illumina Novaseq6000-based sequencing.

Enabling highly scalable PacBio HiFi™ target Capture assays with multiplexed long fragment transposition

Maura Costello, Christianto Putra, Ashley Silvia, David Bays, Zac Zwiko, Jessica Smith, Gavin Rush, and Joe Mellor

seqWell, Inc., Research & Development, Beverly, MA, USA

Recent improvements to the scale and data output of long-read sequencing platforms – for example, the PacBio Revio -- have significantly lowered the costs of long-read sequencing. However, for targeted capture assays where pooling large numbers of samples per Revio cell is more economical, significant challenges still remain in being able to efficiently prepare and multiplex those libraries at scale.

Major barriers include the mechanical fragmentation and adapter ligation methods currently utilized. For example, the Diagenode Megaruptor 3 instrument, which performs DNA fragmentation via syringe-based hydrostatic shearing, takes approximately 1-2 hours per maximum batch size of 8 samples depending on desired fragment size. Fragmenting 96 samples can take multiple days using a single instrument.

To address these challenges, we have developed a multiplexed transposase-based library workflow that allows for efficient and rapid instrument-free generation of long fragment (~8 kb) libraries barcoded with unique dual indexes which can be pooled for hybrid capture and subsequent PacBio sequencing. The entire library workflow takes ~3.5 hours and is done in an easy to automate plate-based format, potentially saving multiple days' worth of time and labor associated with mechanical shearing and adapter ligation workflows per batch of 96.

In this study, we evaluate the application of this new long read tagging method for hybrid capture targeted sequencing using a commercially available probe set relevant to pharmacogenetic gene targets, and compare to data generated using mechanically sheared ligation libraries. The results of the study demonstrate improved uniformity of target coverage (Fold80 base penalty), equivalent SNP concordance in known control samples, and improved reproducibility of fragment size and library yield. We also show that this new transposase-based multiplexing approach can support flexible batching of up to 96 or more samples on a PacBio Revio SMRT-cell, opening new horizons for scalable, high-performance and cost-effective long-read targeted sequencing workflows.

Enhanced single cell DNA methylation analysis using combinatorial sequencing

Maggie Nakamoto, Sanika Khare, Hosu Sin, Ashley Woodfin, Eric Pu, Lin Lin, Phuong Dang, Dominic Skinner, Aimee Beck, Beth Walczak

The epigenetic landscape of the human brain undergoes a plethora of changes during natural development and in malignant transformation. In recent years, DNA methylation-based epigenetic modifications have been widely studied using traditional techniques like bisulfite sequencing and enzymatic methyl sequencing (EM-seq). However, these methods analyze bulk cell populations and lack the granularity of single-cell analysis. Although advances in single-cell analysis have revolutionized our understanding of cellular heterogeneity and functional diversity within complex biological systems, they are still costly, low throughput, and laborious.

To address these challenges, ScaleBio has pioneered combinatorial indexing technology, enabling a significant increase in cell throughput. This method utilizes the cell itself as a compartment to perform 2-3 rounds of sequential barcoding in a plate-based workflow, eliminating the need for complex instrumentation. This technology has been successfully adapted to assess DNA methylation at the single-cell level offering a robust, affordable, high-throughput protocol that enhances yield, diversity, and coverage.

In this study we used ScaleBio's single-cell methylation kit to investigate DNA methylation patterns during development and oncogenesis using fetal and adult brain cells along with cancer cells with widespread DNA methylation changes, such as isocitrate dehydrogenase (IDH) mutant glioma cells at the single-cell level. The IDH gene family, comprising of IDH1, IDH2, and IDH3, encodes enzymes involved in the tricarboxylic acid (TCA) cycle. Mutations in IDH1/2 genes are frequently found in tumors and exhibit a distinct hypermethylation pattern.

(continue to page 197)

(continued from page 196)

We achieved high cell recovery and robust cytosine coverage throughout our analysis of single cell methylomes isolated from brain tissue. Using this data we generated a ranked list of the top hypo- and hypermethylated genomic regions and identified cell type specific clusters seen in different developmental or pathological states by looking at Differentially Methylated Regions (DMR), and compared them to known bulk methylation profiles, uncovering unique single-cell methylation profiles that may be obscured by bulk or pseudo-bulk analysis.

These data show that the ScaleBio single-cell methylation workflow offers increased sensitivity, specificity, and accuracy in identifying DNA methylation sites when compared to other techniques while offering a comprehensive view of methylomes and providing insights into cellular heterogeneity and trajectories.

Enrichment of viable cells from patient samples with LeviCell technology as input into the ResolveOME single cell multiomic workflow

Alison Rojas, Swetha D. Velivela, Jessica N. Borja, Jesse A. Ramirez, Tatiana V. Morozova, Jeff G. Blackinton, Jon S. Zawistowski

BioSkryb Genomics, Durham, NC.

Single cell multiomic analysis is at the forefront of studies driving insight into tumor clonal evolution and somatic mosaicism in normal tissues, as well as into edited cell line surveillance. The ability to generate unified genomic and transcriptomic information with fidelity is wholly dependent on the input of viable cells into the amplification chemistry, otherwise concurrent genotype and phenotype are not accurately captured. We present the coupling of the Levitas LeviCell 1.0 instrument, where viable vs dying/dead cells differentially levitate in a microfluidic cartridge within a magnetic field, to single cell dispensing technologies providing input into ResolveOME multiomic amplification technology. Primary cryopreserved bronchoalveolar lavage cells were either directly inputted into FACS or HP D100 instruments for dispensing into 96 well PCR plates or inputted into the LeviCell 1.0 prior to dispensing. For FACS dispensing, even with the use of Calcein-AM and propidium iodide dual viability staining, levitation of primary cells (2 independent samples, 2 replicates each) prior to dispensing increased the percentage of cells yielding genomic amplification product from 51.25% to 88.75% and, for transcriptomic yield, from 52.50% to 97.50%. A similar trend of reduced reaction yield dropouts was obtained for HP D100 dispensing for both genomic amplification and cDNA yield, albeit with a more modest mean improvement of 17%--yet LeviCell viability enrichment importantly reduced the amount of sporadic clogging of the HP D100 microfluidic cartridge. Low-coverage genomic sequencing metrics assessing library complexity and amplification uniformity were overall comparable between pre- and post-levitation, as were the transcriptomic sequencing metrics of mitochondrial transcript percentage, exonic:intergenic read ratio, and number of expressed genes detected. The marked reduction of reaction failures or sub-optimal reaction yields while maintaining robust sequencing metrics with this LeviCell-dispensing-ResolveOME workflow is a facile solution for maximizing the utility of rare and valuable samples, as well as for maximizing the success rate of a given ResolveOME run with often challenging clinical samples.

Evaluating single-cell platforms for high-quality isoform sequencing

James Webber, Daniel Bartlett, Ghamdan Al-Eryani, Christophe Georgescu, Akanksha Khorgade, Victoria Popic, Aziz Al'Khafaji

BioSkryb Genomics, Durham, NC.

Long-read transcriptomics has moved into the mainstream, driven by transformational advances in long-read sequencing platforms, library construction methods, and computational tools. Increases in sequencing throughput have brought into focus the many types of artifacts introduced during single-cell cDNA synthesis. These artifacts complicate isoform identification and quantification, and can interfere with the detection of gene fusions. To determine the impact of single-cell platform on artifact formation we evaluated PIPseq V4 (Fluent Biosciences), 10x 3’ V3.1, and 10x 5’ V2.1 (10x Genomics) for their ability to make high-quality full-length cDNA libraries for isoform sequencing. We prepared single-cell cDNA libraries from PBMCs with each method and found that 10x 3' and 5' short-read libraries yielded more UMIs per cell compared to PIPseq but with a higher cost per cell overall. We then prepared Iso-seq and MAS-seq libraries from the full-length cDNA and sequenced them on the PacBio Revio sequencer. Our analysis shows that PIPseq produced the fewest artifacts during cDNA creation, the highest proportion of full-splice matches to the reference transcriptome, and the lowest proportion of intergenic and antisense reads. The nominal rate of artifact formation in PIPseq eliminated the need for a purification step and led to longer cDNA molecules overall. In addition, we show that artifactual cDNA molecules can have an impact on quantification in both the short and long read contexts. These results indicate that PIPseq is an effective platform for high-quality isoform sequencing from single cells.

Expanded Capabilities of CellScape™ for Precise Spatial Multiplexing

Technology & Application Development

POSTER 562

Thomas Campbell[1], Oliver Braubach[1], Karen Kwarta[1], Kevin Gamber[1]

Canopy Biosciences, Saint Louis, MO

Understanding the in situ spatial arrangement of key phenotypic populations is critical in advancing our understanding of cancer to inform the development of novel therapeutics and diagnostics. The recent advancement of several tools and platforms that enable high-plex detection of transcripts or proteins on in-tact tissue specimen has given rise to the exciting field of spatial biology. But many novel platforms face challenges with data quality, robust assay performance, and slow data acquisition that limits their practical utility. The CellScapeTM platform enables high-plex spatial omics with quantitative readouts and streamlined assay design and customization. Here, we present the development and deployment of novel assays to expand the capabilities of the CellScape platform. Multiplex assays were analyzed on a variety of human tumor samples for quantitative single-cell analysis of nearly two million single cells in in-tact tissue, preserving spatial information. Using novel proprietary HDR image acquisition for 8-log dynamic range detection, combined with best-in-class resolution allows for quantitative true phenotyping of every cell in the sample, enabling unparalleled analysis of spatial neighborhoods. Leveraging open-source probe chemistry, robust and reliable assays are developed rapidly, leading to practical adoption and analysis of a wide range or targets.

Expanding the Toolbox for Methylome Analysis through the Discovery of Novel Cytosine Deaminases

Romualdas Vaisvila[1*], Sean R. Johnson[1*], Bo Yan[1], Nan Dai[1], Lixin Chen[1], Billal M. Bourkia[1], Ivan R. Corrêa Jr.[1], Erbay Yigit[1], Thomas C Evans Jr[1], and **Zhiyi Sun[1]**

New England Biolabs Inc., 240 County Road, Ipswich, MA 01938, United States
* These authors contributed equally

Cytosine deaminases are very useful enzymes for the detection of epigenetic modifications and for genome editing. However, the range of applications of deaminases is limited by a small number of well-characterized enzymes. To accelerate the discovery and expand the deaminases toolkit, we developed an in vitro screening system that circumvents a major hurdle with their severe toxicity to expression hosts and enables their direct and quantitative characterization. Our results from more than 200 enzymes from 13 deaminase families revealed an unprecedented diversity of substrate selectivity. The discovery of cytosine deaminases with unique properties opens new avenues for biotechnological and medical applications. Using a novel non-specific, modification-sensitive double-stranded DNA deaminase, we developed a single-enzyme methylation sequencing (SEM-seq) method for 5-methylcytosine detection. SEM-seq generated accurate base-resolution maps of 5-methylcytosine in human genome samples including cell-free DNA and ultra-low input DNA (10 pg) equivalent to the amount in a single cell. This streamlined protocol will greatly facilitate high-throughput epigenome profiling of scarce biological material.

Flexible, high performance methylation workflow for challenging samples

Technology & Application Development

POSTER 564

Ushati Das Chakravarty, Karl Spork, Katelyn Larkin, Jessica Sheu, Shengyao Chen, Steve Groenewold, Laurie Kurihara, Steven Henck

Integrated DNA Technologies, Coralville, IA, USA

Epigenetic changes in global and focal methylation of CpG islands are widespread and have been shown to be important biomarkers of a variety of disease states, including cancer. In clinically relevant samples, which are often limited in quantity, purity, and quality; targeted epigenetic sequencing approaches coupled with highly efficient library preparation can provide a higher-resolution and cheaper alternative to whole genome methylation sequencing. Here, we present a future proof methylation workflow which provides the user the flexibility to use either bisulfite or enzymatic conversion upstream of single stranded library preparation for highly efficient and template independent adapter attachment. This method overcomes both disadvantages of dsDNA methylation library prep: it reduces 3' methylation artefacts and maintains high library complexity from low input samples, regardless of conversion module.

Traditionally, post-conversion single-stranded DNA (ssDNA) based library construction has been employed for bisulfite methylation sequencing; primarily because dsDNA-based library preparation followed by bisulfite conversion results in decreased library complexity due to bisulfite induced library fragmentation. Conversely, more mild enzymatic conversion methods are currently enabled by pre-conversion dsDNA-based library preparation. However, such methods require end polishing for dsDNA adapter ligation which introduces synthetic cytosine methylation artefacts, up to 25 bases in length, at aligned 3' ends when 5' overhangs are filled in with unmethylated cytosines. In this study we demonstrate the benefits of TET2/T4-βGT and APOBEC3A enzymatic conversion for use upstream of xGen Methyl-Seq, a ssDNA library construction approach. Whole genome methylation sequencing resulted in greater mapping efficiency, CpG coverage and library complexity, when compared to standard dsDNA library preparation with downstream enzymatic conversion. To further reduce the cost of sequencing, a targeted custom methylation design strategy was employed to capture a broad range of methylation states with high efficiency. The methylation signatures from targeted methylation sequencing are shown to be highly correlated with whole genome methylation results demonstrating the improved performance of xGen ssDNA methylation libraries and the benefits of targeted methylation capture.

Full-length isoform concatenation sequencing to resolve cancer transcriptome complexity

Saranga Wijeratne[1*], Maria E. Hernandez Gonzalez[1*], Kelli Roach[1], Katherine E. Miller[1,2], Kathleen M. Schieffer[1,2,3], James R. Fitch[1], Jeffrey Leonard[4,5], Peter White[1,2], Benjamin J. Kelly[1], Catherine E. Cottrell[1,2,3], Elaine R. Mardis[1,2,5] Richard K. Wilson[1,2], Anthony R. Miller[1]
*Equal contributors

New England Biolabs Inc., 240 County Road, Ipswich, MA 01938, United States
* These authors contributed equally

1 The Steve and Cindy Rasmussen Institute for Genomic Medicine, Abigail Wexner Research Institute at Nationwide Children's Hospital, 575 Children's Crossroad, Columbus, OH, 43215, USA
2 Department of Pediatrics, The Ohio State University College of Medicine, Columbus, OH, USA
3 Department of Pathology, The Ohio State University College of Medicine, Columbus, OH, USA
4 Department of Neurosurgery, Nationwide Children's Hospital, Columbus, Ohio, USA
5 Department of Neurosurgery, The Ohio State University College of Medicine, Columbus, OH, USA

Cancers exhibit complex transcriptomes with aberrant splicing that induces isoform-level differential expression compared to non-diseased tissues. Transcriptomic profiling using short-read sequencing has utility in providing a cost-effective approach for evaluating isoform expression, although short-read assembly displays limitations in the accurate inference of full-length transcripts. Long-read RNA sequencing (Iso-Seq), using the Pacific Biosciences (PacBio) platform, can overcome such limitations by providing full-length isoform sequence resolution which requires no read assembly and represents native expressed transcripts. A constraint of the Iso-Seq protocol is due to fewer reads output per instrument run, thus impacting the detection of lowly expressed transcripts. To address this deficiency, we developed a concatenation workflow, PacBio Full-Length Isoform Concatemer Sequencing (PB_FLIC-Seq), designed to increase the number of unique, sequenced PacBio long-reads thereby improving detection of moderate to low-level expressed isoforms.

(continue to page 204)

(continued from page 203)

In sequencing a commercial reference (Spike-In RNA Variants; SIRV) with known isoform complexity we demonstrated a 3.4-fold increase in read output per run and improved SIRV recall when using the PB_FLIC-Seq method compared to the same samples processed with a non-concatemer workflow. We applied this protocol to a translational cancer case, also demonstrating the utility of the PB_FLIC-Seq method for identifying differential full-length isoform expression in a pediatric diffuse midline glioma compared to its adjacent non-malignant tissue. Our data analysis revealed increased expression of extracellular matrix (ECM) genes within the tumor sample, including an isoform of the Secreted Protein Acidic and Cysteine Rich (SPARC) gene that was expressed 11,676-fold higher than in the adjacent non-malignant tissue. Finally, by using the PB_FLIC-Seq method, we detected several cancer-specific novel isoforms.

This work describes a concatenation-based methodology for increasing the number of sequenced full-length isoform reads on the PacBio platform, yielding improved discovery of expressed isoforms. We applied this workflow to profile the transcriptome of a pediatric diffuse midline glioma and adjacent non-malignant tissue. Our findings of cancer-specific novel isoform expression further highlight the importance of long-read sequencing for characterization of complex tumor transcriptomes.

Genome rearrangements in CRISPR-Cas9 genome editing of hematopoietic stem cells

Suk See De Ravin, Andy Pang, Siyuan Liu, Juan Manuel Carava, Michelle Ma, Bingfang Xu, Yongmei Zhao, Alex Hastie, Xiaolin Wu

Despite the advantages and new applications owing to CRISPR-Cas9 nuclease directed genome editing, there is a growing appreciation of the risks that double strand breaks in the genome can have on the function and safety of the treated cells. Large deletions, aneuploidies, and other structural chromosomal abnormalities have been detected and associated with CRISPR-Cas9 editing and concomitant cell culture. This study describes a new method for an unbiased assessment of genome integrity on pooled cells, based on optical genome mapping (OGM), a technique currently used as an alternative to cytogenetic methods in constitutional and oncology applications. Genome editing offers great potential for ex-vivo hematopoietic stem/progenitor cell (HSPC) gene editing, including for the treatment of Inborn errors of immunity (IEI) and other blood disorders such as X-linked Severe Combined Immunodeficiency (SCID-X1). When considering the general safety of correcting mutations, different corrective gene editing approaches have advantages and caveats. For example, targeted integration (TI) using exogenous DNA donors following CRISPR-Cas9 nuclease mediated double strand DNA breaks (DSB) via homology-mediated repair (HDR) has demonstrated high TI efficiencies for correction of multiple IEIs. However, DSBs and DNA damage responses (DDR) following DSBs that may be compounded by exposure to DNA donors (DNA or adeno-associated virus (AAV)) potentially impairing HSPC fitness and long-term engraftment. Along with CRISPR-Cas9 cutting and insertion of adeno associated virus (AAV) delivered DNA through homologous recombination, other strategies for genome editing have been performed and assessed with OGM in this study: CRISPR base editing (BE), and CRISPR prime editing (PE). To ensure quality of cells and downstream applications, appropriate genome integrity assessment methods are critical. This study compares these genome editing approaches in pooled cells and shows that CRISPR-Cas9 nuclease induces structural variants at the target site in a small fraction of cells that is not detected when BE and PE are used. Further, the structural variants detected at target sites after CRISPR-Cas9 are not stable and are undetectable after culture for an additional two weeks. OGM is a sensitive genome wide assessment of genome integrity useful for off-target and other structural variant analysis in cell and gene therapy applications.

Genotype imputation pipeline for improved accuracy using blended genome-exome sequencing with a diverse reference panel for large-scale population studies

Michael Gatzen[1], Christopher Kachulis[1], Candace Patterson[1], Matthew Coole[1], Edyta Malolepsza[1], Megan Shand[1], James C Meldrim[1], Matthew DeFelice[1], Niall J Lennon[1]

1 Broad Institute of MIT and Harvard, Cambridge, MA, USA

The development of innovative sequencing technologies is providing the research community unprecedented opportunities for conducting large-scale population studies to understand the genetic underpinnings of disease. Whole genome sequencing (WGS) remains the gold standard for genetic studies but the relatively high cost remains a barrier to the feasibility of many population studies. Whole exome sequencing (WES) is a more affordable option, however, the shortcomings of being blind to significant portions of the genome is prohibitive for certain research questions. Imputation from genotyping arrays is limited by only capturing predefined alleles, resulting in reduced applicability to diverse populations and disease characteristics.

We recently launched blended genome-exome (BGE) sequencing which provides full-depth coverage over the exome and low-pass coverage over the entire genome. The genome portion is a basis for imputation that is superior to arrays while the exome portion provides valuable high-confidence exome calls. We evaluate imputation accuracy by comparing genotypes imputed from BGE data sequenced on the Illumina NovaSeqX to matched WGS data for a diverse cohort. We find that imputation using BGE results in superior accuracy to existing genotyping arrays while allowing for leveraging arbitrary reference panels due to BGE sequencing not being limited to certain genotyping sites. Furthermore, we updated the reference panel to use the 1000 Genomes Project + Human Genome Diversity Project panel provided by gnomAD, which adds additional 828 samples from 54 populations to the 1000G cohort. This panel increases the number of covered sites by 91% as compared to the commonly used 1000 Genomes Project panel. As a result the coverage of sites used for polygenic risk scores of 10 chronic medical conditions implemented by the NHGRI eMERGE network increases from 99.3% to 99.8%. We created a cloud-native imputation pipeline using GLIMPSE2 that is optimized for cost-effective processing of large-scale cohorts.

The combination of BGE data as an input for imputation with an improved, more diverse reference panel significantly improves the accuracy of results as compared with current approaches. Combined with the high-confidence exome calling this method provides a cost-effective and accurate solution for population genetics studies without the need for multiple analysis modalities.

High throughput single cell simultaneous DNA and RNA sequencing identified genotypes and phenotypes of breast cancer cells

Kaile Wang[1,2^], Rui Ye[1,3,4^], Shanshan Bai[1,2], Zhenna Xiao[1,2], Lei Yang[1,2,4], Jianzhuo Li[1,2], Yiyun Lin[1,2,4], Emi Sei[1,2], Steven H. Lin[3], Alastair M. Thompson[5], Savitri Krishnamurthy[6], Nicholas E. Navin[1,2,7]

1 Department of Systems Biology, UT MD Anderson Cancer Center, Houston, Texas, 77030, USA
2 Department of Genetics, UT MD Anderson Cancer Center, Houston, Texas, 77030, USA
3 Department of Radiation Oncology, UT MD Anderson Cancer Center, Houston, Texas, 77030, USA
4 MD Anderson UTHealth Graduate School of Biomedical Sciences, Houston, Texas, 77030, USA
5 Department of Surgery, Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, Texas, 77030, USA
6 Department of Anatomical Pathology, UT MD Anderson Cancer Center, Houston, Texas, USA
7 Department of Bioinformatics, UT MD Anderson Cancer Center, Houston, Texas, 77030, USA

^ These authors contributed equally
# Correspondence author

Understanding the relationship between genotype and phenotype in breast cancer cells has been challenging at single cell resolution, mainly because existing high-throughput methods are limited to measuring a single modality and data must be integrated indirectly via computational methods. To address this challenge, we developed wellDR-seq, a high-throughput single cell method that can simultaneously measure the single cell whole genome and transcriptome directly from thousands of single cells in parallel. Using this method, we profiled 17,427 single cells in 6 different ER+ breast cancer patients with either premalignant disease (Ductal-carcinoma-in-situ) or invasive ductal carcinoma (IDC). From these data we identified the epithelial cell-of-origin in 3 cases, showing that ER+ breast cancer cells originated from Luminal Hormone Responsive (LumHR) in the normal breast tissues. We also found that autosomal somatic copy number aberrations were exclusively present in LumSec (< 2%) and LumHR epithelial cells, while chrX gain or loss also occurred in stromal cells (eg fibroblast and endothelial) but a low frequency (<2%) in our datasets. Additionally, our data show the impact of subclonal copy number profiles on gene expression programs, which reflects both gene dosage effects in CNA regions and more complex deregulation in copy-neutral regions. wellDR-seq offers a powerful tool for large-scale single cell genome and transcriptome simultaneous sequencing across different clinical samples, opens new avenue to investigate genotype and phenotypes interactions, identify tumor cells origins, reveal somatic copy number and mutation events, quantify the gene dosage effect and discerning differential genes expression between tumor subclones.

Highly multiplexed protein measurements using SOMAmers with a NGS readout

Technology & Application Development

POSTER 569

Andrew Slatter[1], Carlo Randise-Hinchliff[2], Nithya Subramanian[1], Mike Mehan[3], Fiona Kaper[2]

1 Illumina, Assay R&D, Great Abington, England
2 Illumina, Assay R&D, San Diego, CA, USA
3 Illumina, Bioinformatics & DRAGEN HW SW Dept, San Diego, CA, USA

Measurement of protein abundance in biological samples is crucial for better understanding diseases and therefore health outcomes. High performing protein affinity reagents allow for measurements of thousands of proteins simultaneously. SOMAmers (Slow Off-rate Modified Aptamers) are nucleic acid-based affinity reagents that are generated using SELEX technology and utilise chemically modified nucleotides that facilitate binding to proteins with high specificity and sensitivity. SomaScan is an automated bead-based assay that converts protein abundance into relative abundance of SOMAmers from biological samples. Conventionally, SOMAmers are quantitated by microarray readout on a sample-by-sample basis.

Here, we present a new methodology that converts thousands of SOMAmers into a next-generation sequencing library. The conversion step translates individual SOMAmers into NGS readable barcodes. Individual samples within a batch are further barcoded, pooled, and sequenced on high throughput Illumina platforms for read count analysis. Raw sequencing reads are converted into relative protein abundances through normalization approaches for downstream analyses. The dynamic range of the sample is compressed which provides more efficient utilization of the sequencing flow cell resulting in improved measurement precision. The entire workflow is automated on a TECAN fluent to process blood plasma and serum samples into pooled NGS sequencing libraries.

To determine performance, a comparative analysis of the SomaScan NGS readout to the conventional microarray readout for ~6000 human proteins is presented. The NGS method demonstrates high precision typical of the SomaScan assay. Detection capability of the assay is measured using a healthy reference population using statistical methods. Differentially expressed proteins are detected for known markers for age and sex in a test population, thereby demonstrating functionality of the system for high throughput proteomics.

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

High-plex spatial multi-omic technology at the single-cell level in mouse FFPE brain tissue

Katrina van Raay [1], Alyssa Rosenbloom[1], Brian Birditt[1], Ashley Heck[1], Kimberly Young[1], Tiên Phan-Everson[1], Zachary Lewis[1], Giang Ong[1], Shilah Bonnett[1], Tushar Rane[1], Megan Vandenberg[1], Mithra Korukonda[1], Aster Wardhani[1], Patrick Danaher[1], Trieu My Van[1], Vik Devgan[1], Rudy Van Eijsden[1], Carl Brown[1], Rustem Khafizov[1], David Ruff[1], Margaret Hoang[1], Gary Geiss[1], Caleb Stokes[2], Joseph M Beechem[1]

1 NanoString® Technologies, Seattle, Washington, USA

2 Department of Immunology, University of Washington, Seattle, Washington, USA

Spatially resolved, single-cell transcriptomics and proteomics in mouse neural models reveal neurobiological mechanisms relevant to human disease. However, spatial-omic protocols are analyte-specific and fail to capture multi-omic information within the same single cell. We develop a multi-omic workflow on a Spatial Molecular Imager (SMI), a single-cell spatial biology platform that leverages cyclic in situ hybridization chemistry to enable high-plex detection of proteins and RNAs at subcellular resolution. We measured 68 proteins using the Mouse Neuroscience protein panel (Neural Cell Typing and Alzheimer's Pathology) and 1,000 RNA targets (Mouse Neuroscience RNA panel) on the same 5 µm thick FFPE section of the mouse brain. Our multi-omic workflow demonstrates significant benefits to cell segmentation using cell-type specific markers (GFAP, Iba1, NeuN, in addition to a pan soma marker S6 and nuclear stain DAPI) in neural tissue. As a result, the number of transcripts captured within a single cell increased, most notably transcripts distal to cell bodies. This improved the quality of cell typing within the brain, including the detection of rare cell types and states.

To validate the multi-omic workflow, we profiled the uninfected mouse brain and the brain with West Nile Virus (WNV)-infected encephalitis. As expected from WNV encephalitis, we observed persistent neuroinflammation that includes the recruitment and activation of CD8+ T cells and microglia. We captured the major cell types that comprise inflammatory nodules in post-WNV mouse brain (neurons, microglia, and astrocytes) and identified the signaling pathways that underlie persistent microglial-driven neuroinflammation after WNV encephalitis, which illuminates key aspects of neurodegeneration, neurodevelopment, cell state and signaling, including numerous ligands and receptors involved in neuron-glia communication. Our analysis identified pathways related to inflammation and cellular damage in neurons, astrocytes, microglia and T cells.

(continued from page 209)

Taken together, we used the CosMxTM SMI platform to show, for the first time, that large numbers of mouse neuronal cells can be profiled with both protein and RNA at single-cell resolution in a spatial context. This integrated system maximizes the information content per single cell to enable a mechanistic understanding of infectious disease pathology and inflammatory response in the brain.

High-throughput gene-expression profiling enabled by robotic liquid handling, paper-free laboratory workflow management and automated bioinformatic analysis

Donat Alpar[1], Matthias Haimel[2], Oliver Bergner[1], Armin Weisser3, Nils Janson[3], Thomas Zichner[2], Sergej Nowoshilow[2], Vitaly Sedlyarov[2], Markus Bauer[2], Christoph Disch[3], Onur Kaya[1]

1 Boehringer Ingelheim RCV GmbH & Co KG, Translational Medicine & Clinical Pharmacology, Vienna, Austria
2 Boehringer Ingelheim RCV GmbH & Co KG, Global Computational Biology & Digital Sciences, Vienna, Austria
3 Boehringer Ingelheim Pharma GmbH & Co KG, Global Facilities and Engineering, Biberach, Germany

In oncology research, transcriptome profiling enables genome-wide interrogation of gene expression changes associated with tumor development, progression and resistance. RNA-seq is widely used for target and biomarker identification during drug discovery, aiding preclinical model selection, dose optimization, assessment of pathway modulation and characterization of treatment induced toxicity and resistance. Nevertheless, integration of RNA-seq into workflows built on more conventional methods is commonly hindered by concerns around labor-intensive library preparation, prolonged turn-around time, limited scalability and challenges related to sample traceability and standardized bioinformatic analysis.

Therefore, we have developed a novel RNA-seq workflow by combining i) automated nucleic acid extraction, ii) robot-assisted NGS library preparation, iii) fully digitalized and integrated sample, reagent and workflow tracking, and iv) automated NGS data processing coupled with optimized bioinformatic analysis and interactive visualization.

(continue to page 212)

(continued from page 211)

The manual workflow of the QuantSeq 3' mRNA-seq V2 Library Prep Kit with UDI was converted into an automated protocol requiring 30 minutes of preparation on a Beckman Coulter Biomek i7 liquid handling station. RNA samples extracted from cells/tumor tissues using Maxwell RSC 48 or KingFisher Flex automated nucleic acid extraction platforms were subjected to automated NGS library preparation, commonly executed overnight, with all library pools being sequenced on NextSeq 2000 or NovaSeq 6000 instruments. To support a reliable and highly rationalized laboratory workflow, the bespoke smart app NORBERT (NGS Oncology Research Based Electronic Recording Tool) was developed, which enables a web-based, paper-free laboratory management including fully digital sample, reagent and workflow tracking, with all collected metadata being compliant with FAIR principles and harmonized with upstream (e.g. preclinical in vivo study) and downstream (e.g. omics) systems ensuring the adherence to identical vocabularies and ontologies. For NGS data processing, automated read demultiplexing was established with the output being channeled into a state-of-the-art pipeline directly connected to a proprietary interface which enables data visualization and facilitates tertiary analysis. The event-based cloud infrastructure allows for automated pipeline execution, real-time pipeline monitoring and granular quality tracking. This highly optimized wet-lab/computational approach greatly facilitates the robust and reliable application of high-throughput gene-expression profiling in a wide range of preclinical studies with radically reduced turnaround time.

High-throughput single-microbe genomics enabled by semi-permeable capsules

Technology & Application Development

POSTER 572

Ž. Kapustina, V. Kurmauskaitė, V. Kiseliovas, R. Žilionis

Atrandi Biosciences | Droplet Genomics, Vilnius, Lithuania

Whole-genome and targeted sequencing open a window to understanding the diversity and function of unculturable microorganisms. On one hand, metagenomic sequencing is attractive for its straightforward sequencing library preparation from bulk environmental samples but only offers limited resolution into individual species. On the other hand, single-microbe sequencing offers true single-clone resolution but can only meaningfully address the high biological diversity expected in environmental samples if performed on thousands of individual cells in parallel. To satisfy the need to study such large numbers of single-microbes per sample, well- and droplet-based approaches keep evolving in parallel to provide single-cell compartmentalization required during sequencing library preparation. However, these approaches suffer from a fundamental trade-off between throughput and versatility. Our Semi-Permeable Capsule (SPC) technology combines the throughput of droplets with the versatility of wells by enabling a virtually unlimited number of processing steps on genetic material from thousands of individual microbes in parallel. This study aimed to demonstrate the use of SPCs for barcoding >10,000 individual microbial genomes to obtain single-microbe whole genome sequencing data of unprecedented quality. For proof-of-concept evaluation, we encapsulated well-characterized E. coli and B. subtilis bacteria into SPCs, lysed cells at alkaline conditions (pH 13), amplified their genomes, and employed a split-pool approach to add unique cellular barcodes. Upon sequencing of an aliquot of ~3,000 cells, important technical metrics, such as cross-contamination and genome recovery, were measured to assess the performance of the workflow. The results showed excellent genome retention within SPCs, with <1% of cross-contaminated genomes. Genome recovery analysis yielded a median coverage of 90% at a median sequencing depth of 8X for B. subtilis cells (1690 cells sequenced), and a median coverage of 63% at a median sequencing depth of 3X for E. coli cells (2023 cells sequenced) with clear indication for higher recovery at saturating sequencing. We conclude that the compartmentalization of microbial cells into SPCs allows the generation of high-quality whole-genome data at scale, and further apply the method on environmental samples.

Identification of best library amplification enzymes for NGS

Technology & Application Development

POSTER 574

Michael A. Quail[1], Craig Corton[1], James Uphill[1], Jacqueline Keane[2] and Yong Gu[1]

1 Wellcome Sanger Institute, Hinxton, United Kingdom
2 University of Cambridge, United Kingdom

PCR is a major cause of bias and poor yield in sequencing library prep, often resulting in preferential amplification of shorter and GC neutral fragments. Ten years ago we published a paper identifying kapa HiFi as the best enzyme for Illumina sequencing library amplification as it gave the most even coverage across a range of genomes. Since then several new high-fidelity polymerase mixes have been developed specifically for NGS. We have screened a wide range of these enzymes to identify the best PCR enzymes that are currently available for both short and long read library fragment amplification and identify 3 enzymes QuantaBio RepliQa, Watchmaker Equinox and Takara Ex Premier that give more even genome coverage than Kapa HiFi. Additionally RepliQa was identified as the best enzyme for long fragment amplification amplifying 13 and 20kb fragments of yeast genome without significant size bias, which after PacBio HiFi sequencing gave near complete genome assembly from just 1ng DNA input.

IDT's custom oligos for NGS using MGI systems provide flexible WES workflow options

Technology & Application Development

POSTER 575

Ekaterina Star, Katelyn Larkin, Bosun Min, Lynette A. Lewis, Nick Downey, Longhui Ren, Ellen Schmaljohn, Steve Groenewold, Josh Miller, Ashley Dvorak

The diversity in sequencing platforms enables researchers to investigate the genetic basis of human disease while controlling for data quality and cost per sample. In these studies, samples were converted into NGS libraries using the IDT xGen WES (whole exome sequencing) workflow from library prep to hybridization capture and subsequently sequenced on an MGI DNBSEQ-G400 system. Resulting analysis demonstrated consistently low duplication rate (<5%) and high on-target percentage (~>90%) for exome coverage. IDT's MGI adapters, blockers and oligos synthesis combined with proven hybridization capture technology enables high performance exome studies on MGI systems without utilizing platform library conversion that adds additional PCR steps.

First, IDT's xGen Exome Research Panel v2 was combined with the MGIEasy Universal DNA Library Prep Kit (single and dual index). Second, the IDT xGen DNA EZ Library Prep Kit was customized to be compatible for direct sequencing on MGI platform by leveraging IDT-synthesized custom MGI adapter and dual index primers. Hybridization capture was done using IDT's xGen Hybridization and Wash Kit with IDT's custom blocking oligos and post-capture amplification primers compatible with DNBSEQ. Both experimental conditions were evaluated for performance with single and 12-plex captures, and we found sequencing performance remained consistently high through different multiplexing levels.

Although post capture PCR cycles had to be increased to meet MGI sequencer library yield requirements, libraries maintained high complexity with <5% duplication rate for all samples in both workflows. Dual index libraries achieved high on-target of >91% and uniform depth of coverage with >96% bases covered at >20X and >92% bases covered at >30X. These sets of experiments demonstrate how IDT's library and hybridization capture reagents combined with our MGI custom adapters, blockers, and post-amplification primers provide flexibility to researchers in designing their own workflows to balance data quality and cost per sample.

Impact of live cell enrichment methods for single-cell RNA sequencing

Ignas Masilionis, Brigita Meškauskaitė Urben, Tobias Krause, Meril Takizawa, Catherine Snopkowski, Dana Pe'er, Roshan Sharma and Ronan Chaligne.

1 Single-cell Analytics Innovation Lab, Memorial Sloan Kettering Cancer Center, New York, NY, USA
2 Alan and Sandra Gerry Metastasis and Tumor Ecosystems Center; Memorial Sloan Kettering Cancer Center, New York, NY, USA

Single-cell RNA sequencing (scRNA-seq) has emerged as a powerful tool for investigating cellular diversity and gene expression profiles at the individual cell level. However, the method requires the rapid dissociation of fresh tissue by enzymatic digestion, which often induces cellular stress and leads to biased cell type recovery and altered gene expression. To ensure accurate analysis of scRNA-seq data, it is crucial to generate high-quality single-cell suspensions by effectively removing dead cells, debris, and ambient RNA. Here, we undertake a systematic comparison of five methods used to improve viability and quality of single-cell suspensions: Fluorescence activated cell sorting (FACS), Akadeum Microbubble kit, StemCell EasySepTM kit, Miltenyi Biotec Dead Cell Removal Kit, and Levitas Bio LevicellTM Live Cell Enrichment.

Among the five methods we evaluated, FACS is considered the gold standard technique due to its efficiency and versatility. However, the high-pressure nature of FACS can exert mechanical stress on cells during the sorting process, potentially affecting cellular viability and gene expression profiles. On the other hand, the remaining four methods, while less versatile, utilize milder conditions, which may better preserve fragile cell populations.

In our analysis to date, we used two tissues requiring only no to mild dissociation process: lymph node and cerebrospinal fluid (CSF). Viability of cells was assessed using trypan blue staining after dissociation. The cell suspensions were then subjected to the dead cell removal methods following the manufacturer's instructions. 10,000 cells, from each method, were targeted for sequencing using the 10x Genomics 3' v3.1 kit. The evaluation criteria encompassed hands-on time, mandatory equipment, cost, viability improvement, cell type recovery, gene expression profile, and cellular stress response.

We hope these findings will provide valuable information for researchers to select the most suitable live cell enrichment method based on their experimental requirements, thereby facilitating high-quality scRNA-seq analysis.

Implementing new high-throughput sequencers at scale: early Successes and areas for improvement

Technology & Application Development

POSTER 577

Cole Walsh[1], Andrew Bernier[1], Erin LaRoche[1], Tom Howd[1], Tamara Mason Crisafulli[1], Sheila Fisher[1], Stacey Gabriel[1] and Niall Lennon[1]

1 Broad Clinical Labs, Broad Institute, Cambridge, MA 02141

Two significant new sequencer launches occurred last year - both making possible considerable advancements over the previous generation instrument in terms of cost and throughput. The possibility of lower cost sequencing, at scale, should make it possible to accelerate discovery, create new applications, and ultimately demonstrate increased utility of sequencing in multiple market segments. To fully take advantage of the increased power of the new sequencers at scale, we require adaptations of existing (or development of new) upstream sample processing and downstream data management is required.

We present details of the lab infrastructure and workflow developed to support multiple NovaSeq X Plus instruments and multiple PacBio ReviosTM. The process has so far been demonstrated on five of each instrument, yielding a maximum theoretical capacity of ~8.8 Pb of short read and ~525Tb of long read data annually (assuming an uptime of >=80%).

The NovaSeq X Plus 10B kit represents a 2X improvement in throughput over the previous generation NovaSeq 6000 S4 kit. Fully utilizing the 5 instruments in our lab required new liquid handling scripts for strip tube preparation, new LIMS integrations to track the chain of custody of samples without room for error, and a new application to create sample sheets for on rig DRAGEN activation. Most of these changes were incremental over previous processes.

On the other hand, new scalable processes have been required for the Revio platform because our long read sample volume has not required high scale previously. To fill the capacity of five Revio instruments, capable of producing up to 280 long read genomes at 12X coverage per week, we have integrated four Hamilton Starlet liquid handlers, five Diagenode Megaruptor shearing devices, and multiple Sage Science PippinHT size selection instruments, and have developed workflows that can be accomplished with four lab staff.

Production sequencing is being carried out using these processes and metrics on the success of the implementation will be presented.

Implementing the Agilent Sureselect Clinical Research Exome V4 exome enrichment in diagnostic Whole Exome Sequencing at UMC Utrecht

Bert van der Zwaag[1], Martin Elferink[2], Koen van Gassen[1], Maarten Massink[1], Esther Janson[1], Carola Roelands[1], Hanneke van Deutekom[2], Marcel Nelen[1].

1 Genome diagnostics section, department of Genetics, division Laboratory, Pharmacy and Biomedical Genetics, UMC Utrecht, Utrecht, the Netherlands.
2 Bioinformatics section, department of Genetics, division Laboratory, Pharmacy and Biomedical Genetics, UMC Utrecht, Utrecht, the Netherlands.

The ISO15189 accredited genome diagnostics section of the UMC Utrecht department of Genetics has been offering high quality diagnostic genetic testing through cytogenetic and molecular methods, such as karyotyping, SNP microarray analysis and Sanger sequencing, for over 30 years. Increasingly, genetic testing in our laboratory uses Whole Exome Sequencing (WES) as a basis for its tests. Since the implementation of Next generation sequencing in our laboratory in 2011, a steady growth of this method has seen numbers of WES based analyses grow to well beyond 10.000 annually in 2023. To meet the labs need for a more efficient and comprehensive test, we implemented the novel Agilent SureSelect Clinical Research Exome V4 capture probe set in our routine WES-based diagnostic testing. This new exome enrichment kit offers improved coverage and variant detection in clinically relevant regions in the genome, more even capture allowing better copy-number variant (CNV) detection and higher throughput per sequencing run and the power of 41 'mini genomes' to discover novel and complex variants in frequently mutated genes, like BRCA1 and BRCA2. Using this new probe set for enrichment, routinely delivers exome coverage at >20x depth of >95%, SNV and Indel sensitivity of >99,8% and >98% respectively (GIAB_v2.19), for the coding sequences of the MANE transcript reference set (GRCh37) and the increased evenness of capturing allows for robust CNV detection of variants targeted by >2 probes in the probe set. WES enrichment is performed using 5 MagnisDx prep systems (CE-IVD, Agilent) providing the needed flexibility to run routine and rapid diagnostic flows in parallel. Sequencing is performed to an average depth of ~140x (~18Gb) per sample on NovaSeq6000 and NovaSeq Xplus systems (Illumina) and initial data analysis is performed on site using in-house developed bioinformatic pipelines. Annotation and interpretation are performed using Alissa Interpret genomic data analysis software (Agilent). This results in a streamlined setup enabling us to now perform WES based diagnostics in >10.000 individuals annually (including rapid pre- and perinatal cases) with higher quality WES data and improved odds of detecting clinically relevant variants.

Implementing the Element Biosciences AVITI sequencing system in a genomics core facility operating the Illumina NovaSeq 6000.

E. Beaudoing, K. Bojkowska, S. Calderon, A. Charpagne, M. Dupasquier, C. Peter, V. Praz, H. Richter, R. Sermier, J. Weber, and J. Marquis

1 Broad Clinical Labs, Broad Institute, Cambridge, MA 02141

Before acquiring the Element Biosciences AVITI sequencing system, we wanted to evaluate how the data generated differ from our current NovaSeq 6000, and how difficult will be the co-existence of both technologies within our local genomics core facility.

Using the three most common applications we process, namely single cell RNAseq, bulk RNAseq and Exome-seq, we performed a direct comparison between the Element Biosciences Aviti and the standard Illumina Novaseq 6000 sequencers.

For 10X genomics single cell libraries, clustering analysis did not reveal major differences between the two technologies. By mixing sequencing data obtained from each sequencer at a 50% ratio, we could actually validate that cells with identical 10X cell barcodes cluster together when visualized in a UMAP plot. Minor deviations start occurring at less even mixture ratios. To note, the better sequencing quality obtained with Aviti resulted in a slight increase in the number of cells detected.

Bulk RNAseq datasets were also extremely similar. The main deviation was found for a few low expressed pseudogenes, again probably resulting from the better mapping of higher quality Aviti sequencing reads. Overall, no major changes were observed when performing differential expression analysis.

Subtle sequencing data quality differences can have large effects on variant calling. Accordingly, variant detection analysis performed from our exome sequencing datasets revealed a significant number of distinct SNVs, while insertion and deletion rates remain relatively similar.

As expected, although genome analysis applications will benefit from the Aviti sequencing read quality, technology swap cannot be effortlessly proposed to our core facility users. At the opposite, implementing the Aviti in our sequencing portfolio should be relatively straightforward for gene expression analysis projects started with, or partly requiring the higher throughput of the NovaSeq.

Improved detection of low frequency mutations in ovarian and endometrial cancers by utilizing a highly accurate sequencing platform

Technology & Application Development

POSTER 581

Jiannis Ragoussis

Ovarian and endometrial cancers come within the top-4 for incident cancers as well as deaths in North American women. Cure rates have not improved in 30 years as high-grade subtypes continue to be diagnosed in Stage III/IV. Attempts at early diagnosis have failed because high-grade cancer cells exfoliate and metastasize while the primary cancer is small and undetectable by existing tests based on imaging and blood-based tumor markers.

DOvEEgene (Detecting Ovarian and Endometrial cancers Early using genomics) is a genomic uterine pap test developed in house to identify somatic mutations in uterine brush samples. A high sensitivity error-reducing capture technology (DOvEEgene-SureSelectHS) utilizing duplex sequencing interrogates 23 genes involved in the development of sporadic and hereditary ovarian and endometrial cancers, as well as 9 microsatellites to identify microsatellite instability (MSI). We apply deep duplex sequencing on uterine and saliva (control) samples to detect somatic mutations at <0.1% VAF, interrogate microsatellite loci for instability and low coverage WGS for copy number analysis. The test can discriminate ovarian and endometrial cancers in peri- and postmenopausal women from benign gynecologic diseases common in that age group. This is important because pathogenic somatic driver mutations are also associated with increasing age and benign disease. DOvEEgene incorporates a deep machine-learning derived classifier that can discriminate the mutational signature of these cancers from benign disease with a sensitivity of 70% and specificity of 100% in a population with high background mutational burden.

Here we tested the PacBio Onso system, a highly accurate sequencing technology from PacBio, based on sequencing by binding (SBB), in order to potentially increase sensitivity while reducing costs by reducing required sequencing depth vs the current NGS standard. We sequenced 30 duplex Illumina sequencing libraries produced using the DovEE assay and compared Onso data in non- duplex sequencing mode as well as duplex sequencing mode to the original duplex sequencing method. We also compare the performance of SBB in detecting microsatellite instability. Here, we present this comparison and highlight the benefits of high accuracy sequencing for the detection of very low frequency (<0.1%) somatic mutations, as well as for sensitive MSI detection.

Improvements in variant calling performance in human genome trios on XLEAP-SBS™ chemistry with onboard DRAGEN™ 4.2 analysis

Technology & Application Development

POSTER 582

Cande Rogert[1], Alix Kwasniewska[2], Ilaria Grizzi[2], Tim Merkel[2], Camilla Columbo[2], Carolyn G. Conant[1]

1 Illumina Inc., Core R&D, San Diego, California, USA
2 Illumina Inc., Core R&D, Cambridge, Cambridgeshire, UK

The conversion to XLEAP-SBS™ Chemistry on NextSeq™ 1000/2000 results in increased data quality and faster turnaround times compared to the previous generation of the system. Notable performance improvements are a reduction in run time for the high throughput flow cells as well as significant error rate reduction and an increase in total output – up to 500 gigabases – with a higher density P4 flow cell. The generally improved data quality coupled with DRAGEN™ 4.2 delivers most accurate and comprehensive coverage of the genome in this system to date. We demonstrate improvements in variant calling in human genomes and exomes as well as system expansion up to 40x human trios.

For Research Use Only. Not for use diagnostic procedures. SP-00395

In situ detection and subcellular localization of 5,000 genes using Xenium Analyzer

Robert Henley[1], Amy Oh[1], Hiroshi Sasaki[1], Zhuoran Ma[1], Jordan Sicherman[1], Ian Fiddes[1], Vijay Kumar[1], Florian Wagner[1], Daniel Riordan[1], Patrick Marks[1]

1 10x Genomics, Pleasanton, CA.

Spatial transcriptomics has emerged as a powerful tool that allows researchers to explore the spatial organization of cell types, cell-cell interactions, and cell states by quantifying and localizing gene expression within intact tissues. The Xenium Analyzer provides an end-to-end solution for researchers to perform spatial analysis with highly sensitive and specific detection of RNA. Fully automated multiplexed decoding and computational analysis occurs on the instrument. To date, commercially available panels of 300-500 targets have been used to probe tissue biology using the Xenium Analyzer, allowing for high-resolution exploration of the underlying biology.

Here, we demonstrate how the RNA multiplexing capabilities of Xenium can be 10x increased to up to 5,000 genes, providing comprehensive in situ gene expression analysis for any tissue. We developed a 5,000-plex gene panel that allows for more in-depth pan-tissue cell typing in all major tissues, analysis of cell signaling pathways and identification of genes that are known to have aberrant expression in disease.

We combined this increased gene plex with advances in cell segmentation, using a cell membrane-based approach to define the cell boundaries.
Using multiple human formalin fixed & paraffin embedded (FFPE) tissues (kidney, liver, skin, pancreas, colon, lung, and brain), we show how this increase in plex and segmentation capability helps to improve the accuracy of cell typing, with a wider range of cell types being identified. We also demonstrate that we can identify and characterize unique biological microenvironments across a broad range of healthy and diseased tissue types.

The advancements that we describe here highlight the flexibility of the Xenium platform, with the ability to use smaller plex panels for targeted, defined studies, and higher plex panels for discovery research. This flexibility provides researchers with multiple options for generating high-resolution spatial gene expression data that are key to furthering our understanding of tissue biology.

Inclusive living subcellular sequencing spotlights physical physiological and human pathological features in osteoimmune diversity

Hiroyuki Okada[1,2,3], Yuta Terui[4], Asuka Terashima[5], Masahide Seki[6], Sanshiro Kanazawa[7], Francesca Gori[2], Taku Saito[3], Yutaka Suzuki[6], Roland Baron[2], Ung-il Chung[1,8], Sakae Tanaka[3], Hironori Hojo[1,8]

1 Center for Disease Biology and Integrative Medicine, Graduate School of Medicine, the University of Tokyo; Bunkyo-ku, Tokyo, 113-8655, Japan.
2 Department of Oral Medicine, Infection, and Immunity, Harvard School of Dental Medicine, Boston, MA 02115, USA
3 Department of Orthopaedic Surgery, the University of Tokyo
4 Division of Single Cell Solution, Product Strategy Department, Marketing Center, Life Business HQ, Yokogawa Electric Corporation
5 Bone and Cartilage Regenerative Medicine, the University of Tokyo Hospital
6 Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, the University of Tokyo
7 Department of Oral and Maxillofacial Surgery, Graduate School of Medicine, the University of Tokyo
8 Department of Bioengineering, Graduate School of Engineering, the University of Tokyo

Recent single-cell transcriptomic studies have clarified the cellular diversity within cell populations. Spatial transcriptomics using fixed tissues connects transcriptomics with positional information on histological specimens. However, these technologies cannot decode the dynamic localization of mRNA inside living cells.

To decode intracellular gene expression in living cells, we developed intra-single cell sequencing (iSCseq), a new approach combining confocal imaging, picking up any cellular and subcellular components inside both osteoclasts (OCs) and ordinary small cells using the Single CellomeTM System SS2000 (Yokogawa, Japan), and SMART-seq single cell library for next-generation sequencing.

This study aimed to establish a new technology, iSCseq, to examine subcellular transcriptomic heterogeneity in living OCs and small cells. Matching the data points obtained by iSCseq with conventional scRNA-seq datasets has clarified the physiological and pathological mechanisms of bone resorption in humans.

(continue to page 224)

(continued from page 223)

The number of genes detected with a 3 μm probe was equivalent to that of genes with conventional fluid-based chromium (10x genomics, U.S.), and genes with a 10 μm probe were more than twice those detected with chromium. iSCseq demonstrated subcellular heterogeneity of gene expression in a single cell. Multinucleated cells showed multiple stages of differentiation within a cell. The physical responses and physiological distribution of mitochondria and calcium can be explained by gene expression. Inclusive iSCseq integrated with similar scRNA-seq datasets has elucidated the physiological mysteries in the context of osteoimmune diversity, including hematopoietic stem and mesenchymal cells. Furthermore, we clarified the human skeletal pathogenesis of bone resorption in the tumor microenvironment.
The iSCseq approach has the potential to enhance cellular physiology and provide clues for identifying therapeutic targets for human diseases. (Ref. of iSCseq; Okada, bioRxiv 2022)

Increasing the complexity and quality of total RNA sequencing at scale

Technology & Application Development

POSTER 585

Nicole Francis[1], Jon Bezney[2], Can Kockan[1], Dan Goodman[1], Jason Lam[1], Stephen Montgomery[2], Stacey Gabriel[1] and Niall Lennon[1].

1 Broad Clinical Labs, Broad Institute, Cambridge, MA 02141
2 Stanford University, Stanford, CA 94305

Bulk RNA sequencing has been largely centered on generating and analyzing mRNA libraries, as it is an efficient and cost-effective approach to studying expression in coding regions. However, total RNA methods can provide a more holistic view of the transcriptome. This can include insight into non-coding regions with roles in regulatory mechanisms, and the detection of alternative splicing events and post-transcriptional modifications that can provide us with a deeper understanding of gene regulation and isoform diversity. Historically, rRNA and globin depletion methods of generating total RNA data have faced library quality and scalability challenges. However, Watchmaker Genomics now offers an easy to use, highly scalable total RNA kit with depletion that drastically reduces lab processing time when compared to a traditional poly(A) selection method.

We will present in detail a full comparison of sequencing results between libraries generated with the Watchmaker Genomics Polaris Depletion & RNA LC Kit and libraries generated with a globin depletion & poly(A) selection method. As this was a new library construction method our objectives were to 1) evaluate the quantity and quality of library generated as compared to industry-standard poly(A) mRNA methods, 2) compare standard RNA sequencing metrics from this new workflow to the current RNA workflows used by the Broad Institute's Genomics Platform, 3) assess transcript counts and protein coding genes identified within the total RNA workflow, and 4) determine if the assay was suitable for automation and was therefore scalable. Overall, the Watchmaker total RNA workflow with depletion yielded 120% more complex libraries at higher quantities, 10x more on average, and with 121% lower duplication rates for the same samples. The workflow also identified 15% more transcripts and 3% more protein coding genes than the standard poly(A) method. Broad Clinical Labs will begin offering this assay at scale to provide researchers with RNA data that gives more insight into the whole transcriptome.

Independently controlled wells (iconPCR) enable optimal amplification cycles and integrated sample quantitation and quality assessment to increase NGS sample throughput.

Technology & Application Development

POSTER 586

Delsy Martinez[1], Eric Chow[1], Yann Jouvenot[2], Chris Streck[2], Pranav Patel[2]

1 Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, CA, USA

2 n6 Tec, Inc. Pleasanton, CA, USA

NGS has revolutionized genomics research. From early short read counting applications through today's production-scale sequencing platforms, the research world has seen explosive growth in the data quality and output. Scaling sample prep to match increasing sequencer output has presented challenges in library prep workflows. These challenges include sample qualification, normalization and pooling. To date most labs still rely on the traditional brute force method of individual sample clean-up and quantification prior to pooling. For larger sample sizes, this requires expensive liquid handling robotics, laborious manual manipulation, or costly desktop sequencing runs solely for quantitation. Additionally, many samples are rejected due to under or over amplification that results in artifacts such as bubble products and PCR duplicates which can severely impact both sequencing quality. Low input assay types, such as single cell RNA, ChIP-seq, FFPE, cell-free DNA, often are the most challenging and can require a qPCR run of an aliquot to determine optimal PCR cycles, costing time and precious sample.

Here, we demonstrate a new workflow enabled by the n6Tec iconPCR which couples real-time monitoring with independent thermal control over each well. The number of PCR cycles for each sample is automatically adjusted by the instrument using data from real time measurements on each well. This allows optimal amplification cycles so that each sample reaches the target final concentration without any user interaction and prevents over- or under-amplification, while limiting adapter dimers. The real time data provides two additional benefits. First, final quantification results generated from each well provides information to generate equimolar sample pools without an additional downstream quantitation. Second, the ability to generate melt curves for each sample allows the quantification of the library separately from adapter dimers that impact other quantitation methods. We demonstrate in this work the ability to work with a variety of library types to generate optimally amplified libraries and to pool them without additional QC. This novel method speeds up library prep by eliminating final individual sample cleanup and quantitation, is agnostic to library prep chemistry, and will have tremendous savings in reagent costs, labor and overhead expenses, while using existing consumables.

In-depth characterization of human regulatory T cells using single-cell CITE-Seq and high parameter flow cytometry

Xiaoshan Shi[1], Ali Rezvan[1], Moen Sen[1], Cynthia Sakofsky[1], Gracia Genero[1], Stephanie J. Widmann[2] and Aaron J. Tyznik[2]

1 BD Biosciences, 2350 Qume Drive, San Jose, CA 95131
2 BD Biosciences, 10975 Torreyana Rd, San Diego, CA 92121

T regulatory (Treg) cells are a critical subpopulation of CD4 T cells that suppress immune response and maintain peripheral homeostasis and immunological tolerance. Imbalanced Treg frequency and dysregulation of Treg function may lead to development of chronic inflammatory diseases and cancer. Knowing the crucial role of Treg in immune tolerance, Treg modulation has become an important therapeutical approach to treat immune disorders and prevent transplantation rejection in clinical settings. Despite an increased numbers of studies on the function of Treg, their heterogeneity and mechanism of differentiation remain unclear. Here, we utilized a combination of single-cell CITE-Seq and high parameter flow cytometry to deeply resolve the heterogeneity of Tregs. Taking advantage of magnetic cell enrichment in conjunction with fluorescence-activated cell sorting, the rare Treg population was isolated and processed for downstream single-cell CITE-Seq. By using more than a hundred antibody-oligo conjugates, we identified different subsets of Tregs, uncovered the proteomics signature associated with different subpopulations and revealed donor-to-donor variability in human peripheral Treg compartment. The protein expression pattern and antigen density on Treg revealed by CITE-Seq correlated to that of high dimensional flow cytometry. Trajectory analysis segregates Treg subsets into distinct differentiation pathways with different proteomic signatures ranging from naïve, activated to memory Tregs. These findings improve our understanding of the heterogeneity and differentiation of Treg cells and provide a single-cell proteomic atlas for dissecting the complex roles of Tregs in health and disease.

Insights into Healthy and Alzheimer's Disease Brains using the MERSCOPE® Platform

Technology & Application Development

POSTER 588

Renchao Chen[1], Yuan Cai[1], Darshan Patel[1], Simran Kaushal[1], Leiam Colbert[1], Jiang He[1]

1 Vizgen, Cambridge, Massachusetts, USA

The neuronal system is vastly complex, consisting of various molecularly distinct cell types arranged into different anatomical structures. While single-cell sequencing has elucidated the cellular heterogeneity of the nervous system, such methodologies require cell dissociation which poses challenges to researching links between the molecular characteristics of analyzed cells with their anatomic and functional properties. The advent of technologies such as Multiplexed Error-Robust Fluorescence in situ Hybridization (MERFISH) has enabled the comprehensive mapping of cellular composition and spatial organization within the neuronal system. Here, we showcase the capabilities of the Vizgen® MERSCOPE® Platform — an end-to-end solution for MERFISH technology — in deciphering the molecular and cellular features of human brain tissue under physiological and pathological conditions. We utilized a 960-plex gene panel to achieve high-resolution cell typing in neuronal and non-neuronal cell types in different brain regions, uncovering the molecular and cellular features underpinning anatomical structures. The integration of molecular and spatial information from the same intact tissue facilitated an in-depth understanding of cell-type-specific signaling and regulatory mechanisms. Furthermore, to research the pathogenesis of Alzheimer's Disease (AD), we performed a spatial multi-omics study concurrently measuring gene expression and protein staining in both normal and AD human brain tissue samples. This strategy enabled the identification of cell-type-specific molecular and cellular adaptations, highlighting a spatially dependent response associated with AD. In summary, the MERSCOPE Platform in concert with a 960-plex gene panel allowed us to better understand the spatial organization and interaction of different cell types in the brain and the complex alterations occurring in neurodegenerative diseases such as AD.

Instrument-free, high-throughput, single-cell gene expression analysis with TempO-LINC sensitively resolves rare cell subtypes and molecular mechanisms of toxicity response.

Dennis Eastburn[1], Kevin White[1], Salvatore Camiolo[1], Gioele Montis[1], Nicole Martin[1], Bruce Seligmann[1]

1 BioSpyder Technologies, Inc., Carlsbad, CA

Single cell gene expression assays are critical tools for identifying functional subtypes of cells, as well as changes within cells resulting from pathology or treatment. Current methods have limitations in one or more key areas: high cost, low sample throughput, providing cell-associated reads that either fail to map to transcripts or map to uninformative genes, reliance on reverse transcription and/or insufficient sensitivity to consistently measure low expressed genes from single cells - preventing measurements of many key biomarkers and molecular pathways. The aforementioned limitations have contributed to most single-cell sequencing studies being restricted to small numbers of samples and have prevented their wider scale adoption in biopharma, clinical, translational and applied markets where large numbers of samples become cost prohibitive.

Here we report on the development and performance of a novel genomics platform, TempO-LINC, that adds cell-identifying molecular barcodes onto high-sensitivity gene expression probes within fixed cells. All probes within the same fixed cell receive an identical barcode, enabling the reconstruction of single-cell gene expression profiles across as few as several hundred cells and up to 100,000 cells per run. TempO-LINC simultaneously processes up to 96 cell samples per run, has a robust protocol for fixing and banking cell samples and has a higher gene detection rate than existing single-cell platforms. We also show that TempO-LINC has an observed doublet rate of less than 0.5% for 12,500 cells. Critically, TempO-LINC reduces sample sequencing costs and although we show the assay accurately profiles the whole transcriptome (23,000 transcripts), it can also be targeted to measure only actionable/informative genes and molecular pathways of interest – dramatically reducing sequencing costs yet further. We applied TempO-LINC to assay PBMCs and a human hepatocyte model for toxicology and demonstrated the ability to correctly identify unique cell states and molecular pathways from these populations. TempO-LINC is a disruptive single-cell technology that is ideal for new large-scale applications/studies using single-cell sequencing across thousands of samples while improving data quality.

Integrating deep learning-driven morphology profiling, cell sorting, and single cell RNASeq reveals sample diversity and enriched sub-populations

Technology & Application Development

POSTER 590

Christian Corona[1], Andreja Jovic[1], Kiran Saini[1], Tiffine Pham[1], Ryan Carelli[1], Julie Kim[1], Vivian Lu[1], Senzeyu Zhang[1], Stephane C. Boutet[1], Maddison Masaeli[1]

1 Deepcell, Inc. Menlo Park, CA, USA

Technological advances have ushered in a new multi-omics era that has provided highly granular information on cell types and functions at single cell resolution. Morpholomics has traditionally been used as a readout of cell identity, state, and function. However, the scalability of morpholomics and the ability to sort cells based on morphology has remained challenging. The Deepcell platform, REM-I, performs high-dimensional morphology analysis and sorting of unlabeled, single cells using deep learning on high-resolution bright-field images captured in microfluidic flow. Morphology analysis is achieved by a self-supervised foundation model, 'Human Foundation Model' (HFM), which extracts and combines deep learning features with morphometrics from single cell images. This model provides the basis for profiling and sorting a wide morphological spectrum of cells at scale.

We processed diverse sample types on the REM-I platform, including beads, polyclonal cell lines, and dissociated biopsies. These samples were analyzed by HFM, enabling morphology-based exploration and quantitation of the resulting bright-field images. HFM revealed the presence of morphologically distinct populations within each sample, even populations that constituted a small percentage of the total (~1%). To demonstrate the sorting capabilities of REM-I, we applied it to a 'control sample' consisting of several fixed cell lines at known proportions ranging from 1% to >90%. We used the platform to enrich each cell line from the 'control sample' with high yield and purity.

We next showed that live cells enriched on REM-I are unperturbed and viable, rendering them amenable to downstream molecular analyses. Live, unlabeled samples were processed on REM-I, which led to the discovery of morphologically-distinct cell populations, based on HFM analysis. These cell populations were sorted, and the retrieved live cells were molecularly analyzed by single cell RNASeq (10x Chromium). Single cell RNASeq demonstrated that each sorted population was enriched for particular cell types or phenotypes, relative to the pre-sorted sample. These results illustrated: 1) the ability to perform downstream molecular analyses on sorted cells from REM-I, 2) the platform can be used to parse a wide array of sample types, and 3) there exists a potential link between morphology, cell identity, and molecular signatures.

Integration of live cell holotomography and single cell spatial transcriptomics to visualize acquired and intrinsic resistance to targeted therapy

Dennis Gong[1,2,3], Yi Cui[4], Justine Johnson[5], Rachel Liu[4], Sarah Murphy[4], Joseph Beechem[4], William L. Hwang[2,3]

1 Massachusetts Institute of Technology, Program in Health Sciences, and Technology, Cambridge, MA
2 Center for Systems Biology, Department of Radiation Oncology, and Center for Cancer Research, Harvard Medical School and Massachusetts General Hospital, Boston, MA
3 Broad Institute of MIT and Harvard, Cambridge, MA;
4 NanoString Technologies, Inc. Seattle, WA;
5 Nanolive SA, Cambridge, MA

Cell morphologies represent an integrated manifestation of many phenotypic features including intrinsic cellular processes, changes in cell state, and interactions with neighboring cells. In cancer, live cell imaging approaches have been used to uncover dynamics of cellular response to perturbations like chemotherapy, directly visualize the invasive and metastatic potential of single cells and decode complex phenomena such as membrane blebbing. However, current methods are typically limited by phototoxicity and genetic engineering, which may impact the underlying biology in unpredictable ways, and exhibit tradeoffs among temporal and spatial resolution, field of view, and molecular detail. To overcome these shortcomings, we have developed a morphotranscriptomic method combining the rich phenotypic readout of live cell holotomography with ultra high-plex spatial transcriptomic analysis for comprehensive analysis of cellular processes and interactions. Specifically, we created a custom plate-slide adaptor that enabled us to perform automated, label-free live cell holotomography with the 3D Cell Explorer 96 Focus (Nanolive) to achieve long term monitoring of cells at 200 nm resolution followed by rapid fixation and single-cell spatial transcriptomics (6000+ mRNA targets) with the CosMx Spatial Molecular Imager (Nanotring). Using this integrated approach, we studied KRAS G12D mutant pancreatic cancer cells resistant to specific inhibitors (i.e., MRTX1133) and observed morphological changes throughout the duration of treatment that correlated with resistance. Leveraging the intrinsic heterogeneity present in asynchronous cells in culture, ultra-high-plex spatial molecular imaging characterized transcriptional profiles (650+ median genes per cell detected above background) associated with distinct morphological phenotypes. Importantly, we observed minimal cell loss during cyclical imaging. Finally, we developed a novel method for morphological trajectory analysis to infer transcriptional states across the treatment period and identified mediators of treatment resistance in human cancer cells. We anticipate that our morphotranscriptomic analysis platform will provide a generalizable approach for capturing a complete evolutionary dissection of treatment resistant cell states and their dynamic morphological manifestations.

Integration of mostly-natural sequencing by synthesis with multiple single cell and spatial assays: 10X Flex, Visium and immune profiling

Technology & Application Development

POSTER 592

Gila Lithwick-Yanai[1], Eti Meiri[1], Tyson Clark[1], Zohar Shipony[1], Nika Iremadze[1], Shlomit Gilad[1], Jackson Brougher[2], Dale George[2], Theodore Price[3], Diana Tavares Ferreira[3], Doron Lipson[1]

1 Deepcell, Inc. Menlo Park, CA, USA

Introduction: The recent development of single-cell and spatial sequencing has greatly impacted our understanding of cellular heterogeneity and function. However, the significant costs have been prohibitive for large-scale studies. Previously, we demonstrated the successful use of mostly-natural sequencing by synthesis, a new cost-effective platform, for gene expression libraries. Here, we demonstrate the integration and performance of the Ultima Genomics sequencing technology for additional 10X single-cell platforms, including Visium (spatial gene expression), Flex (fixed RNA profiling), and Single Cell Immune Profiling (V(D)J libraries).

Methods: 10X libraries were sequenced on a UG100 sequencer. Sequence data was pre-processed to simulate paired-end data and processed with the relevant 10X software: 1) 16 fresh-frozen Visium libraries, from spinal cord and dorsal root ganglion, 2) six Flex libraries: two singleplex, two quadriplex and two 16-plex libraries, and 3) two 5' Immune Profiling BCR libraries. Each library was compared with Illumina sequencing. For the BCR libraries, a protocol based on rolling circle amplification was developed to enable sequencing of the relevant 3' portion of the single-ended reads together with the cell barcode and UMI sequences.

Results: The comparison of the Visium results with the Illumina sequencing results for the same samples displayed highly similar spatial clustering, and a high pseudobulk correlation ($rs > 0.98$). For the Flex libraries, the sequencing metrics were very similar. In addition, the pseudobulk correlation was even more highly correlated with Illumina results than previous 3'/5' gene expression comparisons ($rs > 0.99$), likely attributable to the fact that the targeted nature of the reads results in the same portion of the read being sequenced by both the single-ended (Ultima Genomics) and paired-ended (Illumina) platforms. 10X 5' prepared BCR libraries were ligated into single-stranded circularized molecules using PadLock. Rolling circle amplification was performed followed by PCR to create final libraries. Following read processing, over 90% of the cells were productive.

Conclusions: A comparison of the results of each platform with corresponding Illumina-sequenced libraries demonstrated high concordance. Our study offers a promising perspective on the potential use of cost-effective sequencing across various 10x single-cell platforms, with implications for standardized, large-scale research in single-cell genomics.

Interpretable and context-free deconvolution of multi-scale transcriptomic lung cancer data

Technology & Application Development

POSTER 593

Robert Sebra[1,2,3], Daniel Charytonowicz[1,2], Rachel Brody[4]

1 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.
2 Center for Advanced Genomics Technology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.
3 Black Family Stem Cell Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
4 Department of Pathology, Molecular and Cell Based Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

Lung cancer is one of the leading causes of cancer mortality among men and women in the United States. Despite advances in the understanding of molecular mechanisms of lung cancer over the last 40 years, the 5-year overall survival has seen only marginal improvement in recent decades. Recent studies have confirmed the crucial role of tumor initiating cells (TIC) in promoting drug resistance in cancer. It has become clear that cancer cell populations exhibit significant heterogeneity in gene expression, often cycling between low and high progenitor-like states with a significant degree of lineage plasticity (LP). Considering the degree of phenotypic lineage plasticity exhibited by these enigmatic cells, there is an urgent need for more reliable markers and improved methods for TIC identification. This study highlights the power of integrating approximately 50M single cell data units from over 1500 published studies alongside a fully-integrated foundation model and cellular deconvolution tool called Unicell Deconvolve. We combine this approach with updated transcriptional entropy scoring (pySCE) methods, which we will also discuss in detail. Combined, these approaches offer a higher resolution assessment of lineage plasticity, tumor initiation, and annotated cellular heterogeneity in tumor tissue and the associated microenvironment by the analysis of subsets of bulk, single-cell, and spatial transcriptomics data. In addition to these comprehensive machine learning methods, fueled by the abundance of single-cell data generated in recent years, we will explore the future impact of isoform-resolved single-cell data and in-situ spatial data. As a physical and computational toolbox, these applications extend beyond lung cancer and are applicable to all human systems. In tandem with the discussion focused of interpretable machine learning methods and context-free deconvolution of multi-scale transcriptional data in relation to lung cancer, we will present results describing cellular and transcriptional signatures associated with tumor initiation and progression and we will further provide examples of how these factors impact overall survival outcomes.

Interrogating the dark regions of the genome with targeted long reads

Technology & Application Development

POSTER 594

Steve Bruinsma[1], David Tse[1], Brian Ross[1], John Parker[1], Esther Musgrave-Brown[2], Ryan Lusk[3], Leigh Monahan[1], Denise Raterman[1], Aaron Halpern[3], Feng Chen[4], Fiona Kaper[3]

1 Illumina Inc., Madison, Wisconsin, USA
2 Illumina Inc., Cambridge, Cambridgeshire, UK
3 Illumina Inc., San Diego, California, USA
4 Illumina Inc., Foster City, California, USA

Combining targeted long reads with short read technology yields a more-complete genome with improved variant calling and phasing in regions of interest, while maximizing value and throughput. Illumina Complete Long Read (ICLR) with Enrichment is a novel method that robustly generates accurate, targeted long reads from small amounts of genomic DNA. Here, we demonstrate targeted capture of large (>5 kb) fragments with oligonucleotide probes. We describe selection of a panel to give best-in-class coverage across the protein-coding regions of the human genome, including historically difficult-to-map regions. Additionally, we demonstrate how the use of long fragments for target capture enables greater flexibility in target selection with wider spacing between probes. We show that such designs can be used to improve variant detection in difficult-to-map regions and to phase entire genes. Finally, we show how this approach can be generalized to probe panels of various sizes and content.

Inter-site performance comparison of an end-to-end NGS automation technology for whole-exome sequencing (WES) library prep

Rachel Kasinskas

This study tested a fully-automated, self-contained system for NGS library preparation, target enrichment, pooling, normalization, and quantification. Performance was evaluated across multiple independent laboratories and several instruments, and against manual methods, for preparation of whole-exome libraries using various conditions and input types (NA12878, blood-derived hgDNA, and cell line hgDNA). WES libraries were prepared on a NGS automation technology using NA12878 hgDNA and KAPA HyperExome Probes* (50 and 100 ng inputs, n=24; 5 automated runs) in two different laboratories (A and B), resulting in singleplex and 8-plex library pools. Automated and manual library prep performance of the KAPA HyperCap v3.0 workflow* using KAPA HyperExome Probes (100 ng NA12878 hgDNA; n=24; singleplex) were compared. At sites A and B, additional samples were prepared using human whole blood and laboratory cell lines (100 ng, n=24; 2 automated runs). Variant detection by the automated workflow was evaluated with NA12878 gold standard with 24901 high-confidence SNPs in the KAPA HyperExome target region. Inter-site and inter-instrument reproducibility for site A and B was evaluated against two other sites (seven runs, five instruments). Key sequencing metrics were considered to assess repeatability of results for individual instruments at sites A and B (3 automated baseline runs, 100 ng input). The results were similar for each run: percent reads on-target >86.3% (SD<0.4 across all instruments), mean depth of coverage >54X (SD<1.1), and duplication rates of <5.9% (SD<0.6). Metrics for other automated runs were similar (lower input [50 ng], multiplexed samples [8-plex], and other sample types [blood and cell-line DNA]), as well as for manual preparation. When reproducibility was assessed across six instrument sites, greater than 85% reads on-target was observed across all instruments. This fully automated workflow demonstrated robust, consistent inter-site performance against a manual workflow for WES using the same exome probe panel. Thus, this method can reduce the hands-on time and user-to-user variability common in manual methods resulting in greater laboratory efficiency and increased confidence in the results.

Isolation of Single Cells from FFPE Brain Tissue for Probe-Based Whole Transcriptome Analysis

Jason B. Navarro[1], Elizabeth A. R. Garfinkle[1], Jocelyn M. Lucyshyn[1], Richard K. Wilson[1], Elaine R. Mardis[1], Anthony R. Miller, and Katherine E. Miller[1]

1 Nationwide Children's Hospital, Institute for Genomic Medicine, Columbus, Ohio, USA

Clinical research samples obtained from patient surgeries are often only available to researchers as snap-frozen tissue slices or formalin-fixed paraffin-embedded (FFPE) tissue in blocks. For single cell-based analysis of samples, isolating nuclei into a single nuclei suspension from frozen tissue is a common path for sample processing due to the difficulty of successfully obtaining a single cell suspension from frozen tissue. To assess the feasibility of using FFPE tissue for single cell analysis, here we investigate a liberase-based method of generating a single cell suspension from a deparaffinized FFPE block of brain tissue. Due to mRNA degradation from the fixation process, we used the probe-based Fixed RNA Flex whole transcriptome kit from 10X Genomics to process the sample rather than attempting to capture mRNA directly. We then sequenced and compared to previous single nucleus transcriptomics done on a non-fixed frozen tissue sample from the same tissue section. Utilizing FFPE blocks for single cell transcriptomic work adds another option for sample preparation in situations where available fresh or frozen tissue is not available.

Jumpcode Genomics CRISPR Depletion Coupled with Oxford Nanopore Sequencing of Full-length Transcripts

Technology & Application Development

POSTER 597

Robert S. Fulton[1], and Amy D'Albora[1], Lisa Cook[1], and Catrina Fronick[1]

1 McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO, USA

With the recognition of increased utility of full-length cDNA sequence, and the higher cost/read of long-read technologies vs. short-read applications, we aimed to evaluate an enrichment strategy that would enable assessment of key transcripts while maintaining relative expression levels of non-targeted transcripts through elimination of abundant, less-informative content, while taking advantage of informative full-length transcripts. Having full-length context enables identification of expressed single nucleotide and splice variants in individual transcript context rather than extrapolation from short-read data, adding significant utility to full-length transcript assessments. For single-cell assays, it enables the association of those expressed variants to specific cell barcodes and UMI's providing single cell resolution of that variation.

By utilizing CRISPR guide sets designed to eliminate specific targets, we evaluate the reduction in those specified targets compared to those non-targeted transcripts and the value of the increased representation of unperturbed targets. We utilized 2 independent guide sets as well as the combination of both on bulk cDNA samples as well as cDNA derived from single-cell assays. One guide set targets ribosomal, mitochondrial, ubiquitously expressed transcripts, and non-transcription product artifacts from the single-cell systems, and the other depletes over 4000 highly expressed transcripts enabling enrichment of lesser expressed but potentially informative transcripts, often masked by the more highly expressed genes. Those depletion products as well as unperturbed libraries were sequenced on the Oxford Nanopore Promethion 24, and those results are presented here.

KAPA EvoPrep workflow: A streamlined, high-efficiency library preparation method for sensitive somatic variant detection

Jo-Anne Penkler[1], Albert Abrie[2], Samantha Dockrall[2], Duylinh Nguyen[1]

1 Roche Diagnostics Solutions, Pleasanton, California, USA
2 Roche Diagnostics Solutions, Cape Town, South Africa

Precision healthcare and personalized medicine are enabled by accurate variant detection. In the oncology space, somatic variant detection by Next Generation Sequencing (NGS) has potential application across the entire clinical research journey - from early cancer detection and treatment selection, through to minimal residual disease monitoring. Current NGS library preparation methods have laborious workflows, with many reagent tubes and handling steps. These methods often have high input requirements incompatible with low sample availability typical of clinical research samples (primarily formalin-fixed paraffin-embedded tissues, FFPET, and cfDNA from liquid biopsies); and can introduce errors and artifacts which limit sensitive and specific rare variant detection.

Here we present the KAPA EvoPrep workflow, an innovative library preparation method designed to address these challenges. The KAPA EvoPrep workflow not only maximizes molecule recovery and minimizes error rates but also offers a streamlined workflow, enabling the conversion of extracted DNA samples into sequencing-ready libraries in under 2 hours.

Key features of KAPA EvoPrep workflow include

- A streamlined automation friendly workflow with minimal handling steps
- Compatibility with clinical research samples, including cfDNA and FFPET DNA
- Compatibility with a wide range of DNA input amounts, from 0.1 - 500 ng high quality DNA
- Optional duplex molecular barcoding using KAPA Universal UMI Adapters, to facilitate UMI-based error correction
- Compatibility with KAPA Target Enrichment methods, including KAPA HyperCap and KAPA HyperPETE workflows

(continue to page 239)

(continued from page 238)

We demonstrate the utility of the KAPA EvoPrep workflow for the preparation of libraries from clinically relevant research samples; and we demonstrate sensitive variant detection of mutations from cfDNA and FFPET reference material using both KAPA HyperCap and KAPA HyperPETE workflows.

In conclusion, the KAPA EvoPrep workflow represents a significant advancement in somatic variant detection for oncology research. Its streamlined workflow, especially when combined with the KAPA HyperCap or KAPA HyperPETE workflows, offers enhanced sensitivity, accuracy, and flexibility while requiring minimal input DNA. By enabling the detection of rare and clinically relevant mutations, the KAPA EvoPrep workflow has the potential to advance research relating to cancer diagnostics, patient stratification, and personalized therapies, ultimately contributing to a deeper understanding of cancer genetics and precision oncology.

Liquid biopsy with simultaneous DNA methylation and fragmentomics analysis enabled by direct single stranded DNA ligation

Chunming Ding[1,2,3], Zhengquan Yang[3], Miaomiao Cai[3], Huidong Wang[3], Yu Zheng[4], Shengnan Jin[2,3]

1 Fudan University Cancer Center, Department of Pathology, Shanghai, China
2 Zhejiang Innomed Biomedical Co., Wenzhou, Zhejiang, China
3 Wenzhou Medical University, School of Laboratory Medicine, Wenzhou, Zhejiang, China
4 RGENE Inc., San Francisco, CA, 94107

Circulating tumor DNA (ctDNA) present in body fluids such as plasma has generated great interest due to many potential clinical applications that may improve patient outcome. CtDNA methylation and fragment patterns (fragmentomics) are gaining increasing interest in early cancer screening as they may provide tumor and tissue specificity.

The accurate determination of both DNA methylation and fragmentomics has been problematic due to highly fragmented and often damaged DNA molecules. CfDNA and ctDNA may have protruding 5' and 3' ends, nicks and gaps within a double stranded DNA (dsDNA), or may be simply present in a single stranded DNA (ssDNA) form.

We developed a ssDNA-based library preparation method based on an engineered thermostable ssDNA ligase (Hyperligase) capable of highly efficient and non-biased ligation. We performed two head-to-head comparisons with two commercial products using patient cfDNA samples. In the comparison to a commercial ssDNA library preparation kit (Accel-NGS® Methyl-Seq DNA Library Kit), our method was about 200%-300% higher in unique coverage depth. In addition, our method was free of ambiguous 3' end sequences suffered by the commercial ssDNA kit. In the comparison to a commercial dsDNA library preparation kit (NEBNext® Enzymatic Methyl-seq Kit), our method yielded comparable library conversion. More importantly, we found substantial (up to 50-90%) loss of DNA methylation towards the 3' half of the cfDNA molecules by the commercial dsDNA kit. Such loss of DNA methylation was clearly seen in both M. SssI treated cfDNA (presumably 100% methylated) and native cfDNA, with the pattern of signal loss consistent with the hypothesis that new synthesis of DNA in the repair step may result in loss of DNA methylation. The dsDNA-based method also lost the true 3' sequence signals as protruding 3' sequences were digested.

In summary, we developed a ssDNA-based NGS method for analyzing ctDNA epigenome (DNA methylation and fragmentomics) with high efficiency (low cfDNA input and high library conversion rate). Our method may be useful for liquid biopsies aiming for early cancer screening and diagnosis, minimal residual disease detection and general tumor load analysis.

Massively multiplexable enzymatic synthesis of (L)-form DNA using mirror image chemistry – Design, synthesis and functional demonstration of (L)-form 3'-O-modified dNTPs with (D)-form terminal transferase

Bo-Shun Huang, Paramesh Jangili, Jennie Lee, Jonnadula V. S. Krishna, Dae H. Kim

Dxome Inc., Irvine, CA 92618, USA

Rapid acceleration of digital content and data generation worldwide has pointed to the need for next generation data storage system capable of storing ever increasing amounts of data. DNA has been a promising potential solution due to its massive encoding capacity with physical footprint that is orders of magnitude smaller compared to conventional storage material. (L)-form DNA, also known as 'mirror image' DNA has added benefit of being highly resistant to biodegradation providing durability even in extreme environments that is incompatible with natural (D)-form DNA. While (L)-form DNA can be chemically synthesized, limitation in scaling the synthesis to millions of DNA fragments with this method remains a significant hurdle. Here, we report the development of a new chemistry that enables enzymatic synthesis of (L)-form DNA using mirror image (D)-form terminal transferase enzyme with (L)-3'-O-modified-dNTPs.

Four (L)-form nucleotides (A, C, G, and U), chirally inverted from natural D-form, were modified as reversible terminators by using either an azidomethyl or 2-nitrobenzyl moiety as a 3'-OH capping group. To demonstrate the complete chirally inverted versions of the substrate and enzyme system, we chemically synthesized using solid-phase peptide synthesis (SPPS) and native chemical ligation (NCL), a 43.5-kDa (381 amino acids) mirror image mutant terminal transferase that enables for the first time, base by base synthesis of a completely synthetic (L)-DNA using an enzymatic method. (L)-form DNA synthesis with sequences containing homopolymer regions were produced with digital precision with each base addition reaction time of less than 5 minutes. Furthermore, the enzymatically synthesized (L)-form DNA fragment was subsequently sequenced using the (L)-form 3'-O-modified fluorescent reversible terminators and sequencing set up described in a separate companion abstract.

Successful demonstration shown here of both writing (enzymatic synthesis) and reading (SBS sequencing) of synthetic (L)-form DNA has established the complete end-to-end process of DNA storage and subsequent data readout using the new mirror image chemistry.

Measuring genetics, 5-mC and 5-hmC at single-base resolution

Walraj S. Gosal1, Jens Fullgrabe[1], Michael Wilson[1], Robert Crawford[1], Ermira Lleshi[1], Paula Golder[1], Nick Harding[1], Lisabet Andreasen[1], Michael Hodgson[1], Aurel Negrea[1], Helen Sansom[1], Ankita Singhal[1], Minna Taipale[1], Jean Teyssandier[1], Mengjie Li[1], Yang Liu[1], Alexandra Palmer[1], Nikolay Pchelintsev[1], Lidia Prieto-Lafuente[1], Audrey Vandomme[1], Gary Yalloway[1], Páidí Creed[1]

1 biomodal Ltd, Cambridge, United Kingdom

DNA comprises molecular information stored in genetic and epigenetic bases, both of which are vital to our understanding of biology. In human genomes, an epigenetic modification at the fifth carbon of cytosines bases comprises one fundamental pathway by which genes can be silenced or activated. Methods widely used to detect epigenetic modifications at cytosine bases rely on either deamination of unmodified cytosines to read as thymine or borane reduction of the modified oxidised cytosine bases to read as thymine. As a result, such methods fail to capture common C-to-T mutations, and importantly also fail to distinguish 5-methylcytosine (5mC - mainly found in silenced parts of the genome) from 5-hydroxymethylcytosine (5hmC - enriched in active gene bodies and enhancers). Hence, existing methods are unable to read the complete information stored in our genomes in a single workflow.

duet multiomics solution is a new technology which sequences at single base resolution the complete genetic sequence integrated with modified cytosine from low nanogram amounts of cfDNA. By introducing an important high-fidelity methyltransferase during the workflow, we are able to construct an expanded information table that can be used to deconvolute A,T,C,G, 5-mC and 5-hmC simultaneously in a single read within a DNA molecule. Using synthetic controls and reference DNA samples, we demonstrate that this method can achieve accurate measurement of 5-mC and 5-hmC as well as achieving high accuracy for SNV calling. We then use a mouse embryonic stem cell-line, ES-E14TG2A, to map for the first time a simultaneous reading of the genome and epigenome at high depth. In this data, we show how these epigenetic modifications are segregated across the genome and how a 6-letter (epi-)genome can be used to predict gene expression and chromatin accessibility.

In summary, using our modified approach we demonstrate simultaneous, phased reading of all six genetic and epigenetic bases. This tool provides a more complete picture of the information stored in genomes and will have applications throughout biology and medicine.

Mining the plasma proteome; a novel ultrasensitive and high-plex liquid biopsy platform reveals changes in biologically important low-abundance biomarkers.

Yuling Luo*, Wei Feng, Joanne C. Beer, Qinyu Hao, Ishara S. Ariyapala, Aparna Sahajan, Andrei Komarov, Katie Cha, Mason Moua, Xiaolei Qiu, Xiaomei Xu, Shalaka Deshmukh, Shweta Iyengar, Sean Kim, Thu Yoshimura, Li Wang, Ming Yu, Kate Engel, Lucas Zhen, Wen Xue, Chen-jung Lee, Chan Ho Park, Cheng Peng, Kaiyuan Zhang, Kasun Buddika, Dwight Kuo, Lei Fang, Bingqing Zhang, Steve Chen, Yiyuan Yin, Xiao-Jun Ma

Alamar Biosciences, Inc., 47071 Bayside Parkway, Fremont, CA 94538; *Presenting author

The blood proteome holds substantial potential for advancing precision medicine, but it poses significant analytical challenges due to the lack of technologies capable of detecting the vast majority of plasma proteins present in very low concentrations. Here, we present NUcleic acid Linked Immuno-Sandwich Assay (NULISA™), which employs a dual capture and release mechanism to effectively reduce assay background noise and enhance the sensitivity of the proximity ligation assay by approximately 10,000 times, reaching attomolar level of detection. NULISA employs DNA-linked antibodies to convert immunocomplexes to reporter DNA molecules which can be quantified with quantitative real-time PCR (qPCR) or next-generation sequencing (NGS), enabling both ultra-high sensitivity and high multiplexing capacity. NULISA assays are fully automated on the ARGO™ HT System, a high-throughput precision proteomics platform that automates the entire workflow from sample to qPCR data or to a pooled NGS library.

Two high-plex NULISA panels: a 250-plex inflammation panel and a 120-plex central nervous system (CNS) disease panel were developed and tested on disease cohorts. The inflammation panel has the most comprehensive coverage of cytokines and chemokines and other immune-related proteins in a single panel and demonstrated superior sensitivity in detecting low-abundance proteins with high precision, enabling the detection of difficult-to-detect but biologically important low-abundance biomarkers in the blood of patients with autoimmune diseases and COVID-19. The CNS panel is the largest multiplex panel designed specifically for neurodegenerative diseases encompassing all key hallmarks of Alzheimer's and has demonstrated the potential to identify both established and novel proteins associated with various neurodegenerative diseases that were previously challenging to detect in the blood.

(continue to page 244)

(continued from page 243)

In conclusion, NULISA uniquely combines ultra-high sensitivity and high multiplexing in a single platform, enabling focused in-depth analysis of the blood proteome. Automation on the ARGO HT system provides a high throughput and high precision proteomics platform, which should open doors to the low abundance portion of the proteome and provide new biological insights that may lead to a new generation of biomarkers for early disease detection and monitoring and eventually to novel therapeutic approaches.

Molecular Pixelation: optics-free 3D spatial proteomics of single-cells by sequencing

Technology & Application Development

POSTER 604

Simon Fredrik

The spatial distribution of cell surface proteins governs vital processes of the immune system such as inter-cell communication and mobility. However, fluorescence microscopy has limited scalability in the multiplexing and throughput needed to drive spatial proteomics discoveries at subcellular level. We present Molecular Pixelation, an optics-free DNA-sequence based method for spatial proteomics of single cells using Antibody Oligonucleotide Conjugates (AOCs) and DNA-based nanometer sized molecular pixels. Relative locations of AOCs are inferred by sequentially associating these into local neighborhoods using the sequence-unique DNA-pixels, forming >1,000 spatially connected zones per cell in 3D. DNA-sequencing reads are computationally arranged into spatial proteomics maps of each single-cell for 76 proteins.

T-cell migration enables infiltration into tumors, an essential mechanism of all immune therapies of cancer. Studying immune-cell dynamics using spatial statistics on graph representations of the data, known and novel patterns of protein spatial organization was found in chemokine-stimulated T-cells, showing potential to define cell states by spatial arrangement of proteins. Molecular Pixelation will lead to discoveries in immune cell activity as well as new drug targets for oncology.

Morphology profiling driven by deep learning characterizes functional changes in CRISPR knockout cell lines

Andreja Jovic[1], Linda Hsie[1], Thomas Vollbrecht[1], Christian Corona[1], Jeanette Mei[1], Tiffine Pham[1], Ryan Carelli[1], Maggie Wang[1], Kiran Saini[1], Julie Kim[1], Vivian Lu[1], Nicholas Banovich[1], Daniel Schraivogel[2], Lars M. Steinmetz[2], Stephane C. Boutet[1], Maddison Masaeli[1]

1 Deepcell, Inc, Menlo Park, USA
2 EMBL Heidelberg, Germany

CRISPR/Cas9-based screens have revolutionized functional genomics by causally linking genes to their function and cellular phenotype. Uncovering this link between genetic perturbations and cellular phenotypes can be performed using phenotypic assays or sequencing of gRNA repertoires. However, these approaches are end-point, cost-prohibitive, limited in phenotypic scope, or are generally not scalable. There is a need for orthogonal methods for characterizing phenotypes and isolating perturbed cells in a generalizable and scalable way. The Deepcell platform, REM-I, combines deep learning, microfluidics, and high-resolution imaging of label-free single cells to characterize and sort cells of interest based solely on morphology.

To evaluate the performance of REM-I on CRISPR screens, a library of 11 single gene knockouts (KOs) was generated in three cell lines - HEK 293 (Human Embryonic Kidney cells), Jurkat (T Lymphocyte), and K562 (Myelogenous Leukemia) - and compared to a Cas9-only control cell line. This gene set was selected based on previously reported effects on cell morphology. We found several phenotypic patterns in KOs from genes belonging to similar pathways, suggesting that morphology is robust and orthogonal as a functional readout. Our approach also detected rare morphological phenotypes (~1-5% of the population), which highlighted the sensitivity of the platform. Lastly, the 11 single gene KOs from individual cell lines were combined with corresponding control cells to create a CRISPR KO mix, which was then sorted on our platform. The sorted cells were processed and compared to the pre-sorted population through a combination of scRNASeq (10x Chromium) and Oxford Nanopore sequencing, which revealed enrichment of single gene KOs or groups of KOs belonging to similar pathways.

A future direction will be to process large-scale genetic perturbation screens on REM-I and to sort target cells from these screens. We envision that our platform will provide a cost-effective, scalable, and generalizable framework for linking genes to their corresponding function, and potentially provide a means to infer gene function through morphology analysis. In conclusion, we show that deep learning-driven morphology analysis and sorting is a promising approach for characterizing the impact of genotype on phenotype.

Multiomic genomic mapping with long read sequencing

Technology & Application Development

POSTER 606

Paul Hook[1], Bryan Venters[2], Luke Morina[1], Alison Hickman[2], Jamie Moore[1], Michael-Christopher Keogh[2], Winston Timp[1]

1 Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, 21218;
2 EpiCypher, Durham, NC, 27709

Genomic mapping of protein-DNA interactions is an indispensable tool for understanding the functional role of the genome and the complex gene regulatory network that lies atop the genome. Recent technology advances have moved from chromatin immunopreciptiation (ChIP-seq) methods to more novel immunotagged nuclease (CUT&RUN) and transposase (CUT&TAG) approaches. But traditional short-read sequencing methods have limitations in analyzing single DNA molecules and mapping repetitive or homologous genomic regions. The advent of long-read sequencing (LRS) platforms promises to overcome these challenges and revolutionize genomics.
Here we present CUTANA-LRS, a multiomics assay that leverages single molecule sequencing (PacBio/ONT) to concurrently profile protein-DNA interactions, chromatin accessibility, and native DNA methylation in a single assay. A key innovation is the nondestructive epigenomic mapping approach using a novel DNA methyltransferase fusion protein (pAG-M.EcoGII), which we have optimized and wrapped into a full and robust protocol.

Using this enzyme, we have demonstrated high concordance with conventional short-read methods, including CUT&RUN and ChIP-seq. By analyzing reads that are "painted" with the methyltransferase, we can estimate which molecules are bound in a quantitative fashion, giving us a picture of fraction of bound molecules. These molecules can also be phased, to detect which allele is bound by the protein (and antibody complex), observing allele-specific histone marks and TF binding dynamics. We have coupled this to the latest ultralong sequencing methods with nanopore sequencing to generate long single molecules which can be leveraged to look at epigenetic information along these long (100kb+) stretches. Importantly, by coupling this method to native DNA methylation measurement and chromatin accessibility (nanoNOMe), we can perform a "multi-color" measurement at once, profiling multiple different aspects of the epigenome simultaneously on individual molecules. This will help to reveal the collaborative nature of the epigenome gene regulation, revealing contributions from methylation, chromatin accessibility, and protein-binding dynamics.

Multi-omic genomic mapping with long read sequencing

Bryan Venters

Gene transcription is regulated by the complex interplay between histone post-translational modifications (PTMs), chromatin associated proteins (CAPs), and DNA methylation (DNAme). Mapping their genomic locations and examining the relationships between these chromatin elements is a powerful approach to decipher mechanisms of disease, thereby enabling discovery of novel biomarkers and therapeutics. Leading epigenomic mapping technologies (e.g., ChIP-seq, CUT&RUN) rely upon DNA fragmentation to isolate regions of interest for sequencing on short read platforms (e.g., Illumina). This strategy leads to substantial loss of contextual information regarding the surrounding DNA, precluding the identification of multiple co-occurring epigenomic features on a single DNA molecule. By contrast, long-read sequencing (LRS) platforms are capable of sequencing very long reads from a single molecule (typically >10kb), allowing relationships between features on a single molecule to be used to resolve heterogeneity within mixed populations.

Here we report a robust multi-omic method that leverages LRS to simultaneously profile histone PTMs (or CAPs), DNAme, and parental haplotype in a single assay. This nondestructive, epigenomic mapping approach leverages a novel DNA methyltransferase fusion protein (pAG-M.EcoGII) to label DNA underneath antibody-targeted chromatin features, thereby marking sites of interest while preserving DNA molecules intact for LRS. Inspired by our work with state-of-the-art immunotethering-based approaches (CUT&RUN / CUT&Tag), nuclei are bound to magnetic beads to streamline and automate sample processing. Next, adenosines nearby antibody-targeted chromatin features are methylated with pAG-M.EcoGII, which are then directly read from genomic DNA using Oxford Nanopore Technologies or Pacific Biosciences LRS platforms. To determine the capabilities and limitations of this assay, we tested multiple chromatin targets in various cell lines. Importantly, this method is highly reproducible across biological replicates, and highly concordant with orthogonal SRS assays (e.g., CUT&RUN). Further, we showed that this method is a true multi-omic approach by simultaneously profiling histone PTMs, native DNAme (5mC), and parental single-nucleotide variants from single DNA molecules within a single reaction. Finally, this workflow preserves chromatin integrity for LRS, revealing heterogeneity (e.g., haplotype or paternal origin) within / between data types and providing access to previously unmappable genomic regions (e.g., centromeres).

NEED-seq: targeted in situ sequencing of epigenetic regulators in fixed cells

Heather M. Raimer Young, Pierre-Olivier Esteve, Laura Blum, Sagnik Sen, Matthew Campbell, V. K. Chaithanya Ponnaluri, Julie Beaulieu, Jim Samuelson, Shuang-Yong Xu, Sriharsa Pradhan, Louise Williams

New England Biolabs, Ipswich, MA 01938, USA

Epigenetic regulation of gene expression is increasingly associated with development and disease. The current methods used to investigate histone modifications, transcription factor binding, and other protein-DNA interactions include CUT&Run, CUT&Tag, and ChIP-seq. Each of these methods have their own limitations. For example, CUT&Run and CUT&Tag do not work on fixed cells. Meanwhile, ChIP-seq can be used on fixed samples but it also requires the isolation and fragmentation of chromatin prior to profiling. NEED-seq (Nicking Enzyme Epitope-targeted DNA sequencing) is an alternate method to investigate in situ profiling of protein-DNA interactions in formaldehyde fixed samples.

NEED-seq is a simple process whereby cells are formaldehyde fixed followed by chromatin relaxation. This exposes epitopes even from closed chromatin regions. Primary antibodies then bind to selected epitopes which enables the subsequent targeted nicking of DNA flanking the bound primary antibody. Following decrosslinking of proteins, genomic DNA is then isolated, and Illumina libraries are prepared. Sequencing data accurately identifies the in situ protein-DNA interactions.

NEED-seq data were generated from fixed cell culture samples for open and closed histone modifications, chromatin organizational proteins, and transcription factors. NEED-seq shortened the experimental timeline from cell fixation to final libraries compared to ChIP-seq from four days to two days. IGV tracks demonstrate comparable profiles between NEED-seq and ChIP-seq data. NEED-seq libraries also showed improved library and peak calling metrics. Overall, the ability of NEED-seq to profile protein-DNA interactions from the original chromatin context will improve our understanding of chromatin dynamics and provide a pathway towards efficient profiling in other fixed samples, such as FFPE tissues.

New streamlined hybridization-capture enables flexible inputs and hyb times without hot liquid handling.

Technology & Application Development

POSTER 609

Alexandra Wang, Daniel Saatman, Jinglie Zhou, Anthony Tsai, Michael Salgado, Steven Henck

Integrated DNA Technologies, R&D, Redwood City, CA USA

Targeted sequencing by hybridization-capture is a versatile and powerful way to interrogate selected regions of interest in large and complex genomes. However, the laboratory workflow for this approach is notoriously complex and lengthy for commercially available methods. Current workflows invariably require troublesome steps such as preheating and handling small volumes of hot liquids in a time- and temperature-sensitive manner. These steps are the laboratory pain points that make hybridization-capture cumbersome to execute and challenging to automate. Optimum performance generally necessitates overnight hybridization and high library inputs. Shorter hybridization times on existing platforms negatively impact targeting, coverage, and GC-balance. The high library input mandates additional PCR cycles that can lead to increased duplication rate and can complicate variant calling.

Here we present a new hybridization-capture workflow that eliminates the laboratory pain points. Benchmarked to our current commercial kit, the number of hands-on steps and reagents are reduced by more than half. The new workflow has no pre-heating, no hot-liquid handling, and no temperature-sensitive urgency. The simple user-friendly workflow increases throughput, decreases hands-on time and is configured for easy automation adaptation. The design and formulations enable equivalent targeting and coverage performance for 1-hour compared to overnight hybridization times with as little as 100ng total library input per hybridization-capture.

Here we illustrate the performance of this streamlined workflow with IDT commercial panels AMLv3 (~1.2Mb) and Exome v2 (>30Mb) where 100 ng input at 1-hour hybridization time achieves equivalent or better targeting, coverage, and GC-balance as current IDT xGen Hybridization-capture with 500ng input and overnight hybridization. The flexibility in hybridization time, library input, and easy workflow makes this hybridization-capture adaptable to a large variety of operational and application needs, from quick surveys to high-throughput and/or in-depth interrogations.

NicE-seq: an approach for high resolution profiling of accessible chromatin on formaldehyde-fixed cells

Sanda Zolj, Laura Blum, Pierre-Olivier Estève, Bradley W. Langhorst, Sriharsa Pradhan, Louise Williams, V K Chaithanya Ponnaluri

New England Biolabs Inc., 240 County Road, Ipswich, MA 01938, USA

Genomic features, including promoters, 5'UTRs, 3'UTRs and enhancers play a crucial role in regulating gene expression. Access to these genomic features within chromatin greatly influences the efficacy of regulatory DNA binding proteins, such as transcription factors and chromatin remodelers. Therefore, understanding the dynamics of chromatin accessibility (open vs closed chromatin) is important for interpreting processes that govern cells and/or context-specific gene expression. Traditionally used methods for the interrogation of chromatin accessibility, include DNase hypersensitive assay, FAIRE-seq, NOMe-seq, and ATAC-seq. The most common approach, ATAC-seq, utilizes a hyperactive Tn5 transposase-based tagmentation approach. For unfixed fresh cells or whole tissues, ATAC-seq works well for input ranges between 25,000 to 100,000 while for clinically relevant fixed cells the recommended input range is between 50,000-100,000 cells. Sequencing data from ATAC-seq libraries provides information about chromatin accessibility, nucleosome occupancy and transcription factor footprints. However, a drawback is that ATAC-seq and Fixed Cell ATAC-seq libraries have a significant amount of mitochondrial contamination thus requiring a high sequencing depth (~100 million paired reads) leading to a high duplicate rate and are more costly.

Here, we describe NicE-seq (Nicking Enzyme assisted sequencing) for high-resolution accessible chromatin profiling focusing on 4% formaldehyde-fixed cells which is similar to clinically available FFPE samples. NicE-seq captures and reveals accessible chromatin sites (OCSs) exposing transcription factor occupancy at single nucleotide resolution. NicE-seq libraries can be made from a broad range of cell numbers (1000 to 100,000) with lower sequencing requirements (~40 million paired reads). OCSs identified by NicE-seq coincide with both DNase hypersensitive and ATAC-seq sites. Metrics including FRiP score, replicate overlap, and peak numbers are comparable between Fixed Cell ATAC-seq and NicE-seq. However, the mitochondrial contamination in NicE-seq samples is remarkably lower (5-10%) compared to the 30-50% observed in the Fixed Cell ATAC-seq method. Furthermore, NicE-seq avoids column purification and instead utilizes a bead-based approach enabling a flexible automation-friendly workflow. Taken together, NicE-seq offers an optimized solution for working with a broad range of cell inputs and sample quality (formaldehyde-fixed cells) while maintaining high data quality metrics and improved mitochondrial contamination avoidance. NicE-seq is cost-effective, easy to use, and scalable for automation.

Novel Enzymatic DNA Fragmentation for Long-read Sequencing.

Technology & Application Development

POSTER 612

Brittany S. Sexton, Margaret R. Heider, Louise Williams, Laura Blum, Bradley W. Langhorst, and Pingfang Liu

New England Biolabs, Inc.

DNA fragmentation upstream of long-read sequencing can be used for low input samples and to obtain comparable fragment sizes across samples and experiments. Additionally, it has been shown that fragmentation upstream of long-read sequencing can result in higher N50 lengths, due to very long DNA fragments not being efficiently captured in library preparation. Mechanical shearing is the current method for control of DNA fragmentation size upstream of long-read sequencing library preparation. Mechanical shearing requires costly instruments, can be difficult to automate for high throughput labs, and can cause DNA damage resulting in sample loss. In comparison, enzymatic fragmentation methods do not require expensive instruments and are automation-friendly. However, there are currently no enzymatic fragmentation methods that allow for fine adjustment of DNA fragment size upstream of long-read sequencing.

We have developed a novel enzymatic fragmentation method that can be used upstream of library preparation for long read sequencing. This method is quick and robust, taking as little as 20 minutes. Enzymatic fragmentation is automation-friendly, time-dependent and can be used to generate a wide-range of DNA fragment sizes suitable for different sequencing platforms.

Here we demonstrate the use of the novel enzymatic fragmentation method upstream of Oxford Nanopore Technologies library preparation to generate high quality libraries and comparable N50 lengths to mechanical shearing (Covaris G-tube). This new, robust enzymatic fragmentation method overcomes the limitations of mechanical methods, simplifies sample processing, increases throughput, and improves the library preparation workflows and sequencing quality for long-read sequencing.

One-Step reverse transcription multiplex targeted amplification of human respiratory syncytial virus for whole-genome sequencing.

Kaylinnette Pinet[1], Matthew Angel[1], Anna-Maria Franco Yusti[2], Alexandre Dosbaa[2], Romain Coppée[2], Yanxia Bei[1], Pingfang Liu[1], V K Chaithanya Ponnaluri[1], Bradley W. Langhorst[1], Quentin Le Hingrat[2], Keerthana Krishnan[1]

1 New England Biolabs, Ipswich, MA, USA
2 Bichat-Claude Bernard University Hospital, Paris, France

Human respiratory syncytial virus (RSV) is a single-stranded RNA virus that has persisted for over seven decades as a leading cause of respiratory illness in infants and young children. Applying multiplex targeted amplification sequencing techniques to RSV can support public health and research efforts to monitor vaccine or therapeutic evasion mutations and emergent strains. Here we describe a newly developed RSV sequencing approach that was applied to over 20 clinical samples with a broad range of Ct values (Ct = 19 - 40).

Our amplicon-based targeted sequencing method for RSV uses two multiplexed primer pools that generate overlapping amplicons and optimize genome coverage for both major RSV subtypes, A and B. This streamlined RSV sequencing approach combines reverse transcription and targeted amplification into a single step via a reverse transcriptase and high-fidelity DNA polymerase enzyme mix. Following cDNA synthesis and targeted amplification, libraries can be prepared for Oxford Nanopore Technologies® or Illumina® sequencing platforms. Furthermore, our RSV sequencing method reduces plastic consumption, hands-on time, and the probability of sample-cross contamination. Thus, this workflow provides scientists and public health experts with a cost-effective and efficient solution for the monitoring of RSV viral genomes to ensure and improve vaccine efficacy.

On-flow cell weighted low pass genome sequencing simplifies the workflow and accelerates breeding programs in agriculture

Junhua Zhao[1], Adeline Mah[1], Yirong Yang[2], Laure Edoli[1], Connor Thompson[1], Xin Wu[2], Jianxiang Peng[2], Bryan Lajoie[1], Shawn Levy[1], Benjamin Song[1], Bingbing Wang[2], Michael Previte[1]

1 Element Biosciences, San Diego, CA, 92121
2 Biobin Data Sciences, Changsha, Hunan, China

Low pass genome sequencing along with genotype imputation has been shown as a cost-effective alternative to genotyping arrays for trait mapping in the agriculture field. While research has shown the feasibility of driving down the sequencing depth to 1x or 0.5x, coverage uniformity throughout the genome and low coverage of important trait sites usually make it challenging to increase sample numbers in each sequencing run to further drive down sequencing depth. To resolve this, additional target capture library preparations to enrich the interest sites can provide valuable information when combined with low pass sequencing. However, these additional library preparation processes are usually costly and complex. Recently, we explored the application of a probe-specific flow cell to enable low pass sequencing and selected enrichment simultaneously, using a standard whole genome sequencing library.

We collected tens to hundreds of important single nucleotide polymorphisms (SNP) / Quantitative Trait Nucleotides (QTN) in maize and assigned ranking scores to them. As an initial test, we designed the corresponding probes targeting 11 SNP/QTN sites and produced probe-specific enriched flow cells for this study. The whole genome sequencing library preparation was performed using gDNA extracted from the seeds of maize inbred and hybrid strains (B73, B73xMo17, etc.). The libraries were sequenced on an AVITITM instrument with probe-specific flow cells, without further enrichment process. We down-sampled to 1x coverage and observed whole genome coverage CV below 0.05-0.06, and 40-50x fold enrichment across the 11 probe sites, providing enough coverage to confidently call SNP/QTN. We then down-sampled the sequencing data to 0.5x and 0.3x coverage to apply our imputation algorithm to evaluate concordance between coverage depths.

This study showed this novel technology of probe-specific enriched flow cell achieved uniformed whole genome coverage and simultaneously captured interesting trait SNP/QTN sites with higher coverage to confidently provide genotyping information at much lower low cost and with a simplified workflow. The described workflow bridges the gap between the completeness of sequencing with the cost-effectiveness of microarray approaches and will facilitate the breeding process in many agricultural species.

Optimisation of an automated nanopore genomic sequencing pipeline.

Richard A Moore, Michael Mayo, Pawan Pandoh, Yongjun Zhao, Robin Coope, Duane Smailus, Andrew Mungall, and Steven Jones.

BC Cancer, Michael Smith Genome Sciences Centre, Vancouver, BC Canada.

Long read sequencing of genomes has become the gold standard for many applications including de novo assembly, cancer genomics, infectious agent detection and inherited disease. As demand has increased we have established an automated library construction process utilizing the Hamilton Nimbus, enabling generation of libraries on 96 well plates.

Nanopore sequencing is very dependent on input DNA quality and source, and can therefore result in a range of yields. We have optimized a pipeline that yields consistent results in terms of read lengths and yields, whilst enabling collaborators to submit DNA from various extraction methods and sources.

Size selection upfront was found to be critical, and we have implemented a Blue Pippin based size selection to remove smaller DNA fragments. A gentle shearing using Covaris G-tubes has been added and optimized to maximize yield at the sacrifice of ultra long reads.

Additional improvements have also been tested including run time extension, DNA repair, and additions to running buffer to remove secondary structures and increase read length, this is particularly impactful on samples with repetitive genomes. Nuclease flushes and sample re-loading in the first 48hours optimizes the utilization of the flow cell.

The pipeline as described is now robust to variable inputs yielding long read genomic sequence with an N50 in the 15kb range and 120-150Gb of data from the latest R10 chemistry.

Optimization of DNA library preparation workflows for emerging sequencing platforms

Technology & Application Development

POSTER 616

Pingfang Liu, Chen Song, Margaret R. Heider, Kruti Patel, Matthew Campbell, and Bradley W. Langhorst

New England Biolabs, Ipswich, MA 01982 USA

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, Element Biosciences, Ultima Genomics and Singular Genomics. Lower run costs, faster run times and ease-of-use offered by new sequencing companies renders adoption of emerging platforms an appealing and reasonable choice. However, data quality, error rates, and error types can vary among sequencers, including between instruments from the same company. Therefore, it is important to understand the performance of different sequencing platforms to ensure the continuity of long-term projects and if necessary, to develop mitigation strategies to overcome discrepancies in error profiles.

Here, we present two sample preparation workflows that have been optimized for the Element AVITI and MGI sequencing platforms, respectively. One method utilizes mechanically sheared DNA as the starting material, whereas the other method utilizes a robust, flexible, enzymatic fragmentation method; both workflows are compatible with sequencing chemistries used by both platforms. Using these workflows, we observed high-quality data with superior sequencing metrics, including increased mapping rate, lower duplication, lower chimeras, lower base bias, and increased genome coverage compared to alternative library prep workflows. Additionally, our method can be used to generate libraries for other sequencing platforms, allowing for direct comparisons of different sequencing technologies. We have also developed qPCR-based library quantitation methods for the new platforms, including Ultima, to enable accurate library quantification to maximize sequencing output and data quality. Modular and tunable library preparation approaches plus accurate quantification will further facilitate workflow management and flexible reagent inventory for users working with multiple sequencing platforms.

Performance evaluation of host depletion methods for improved detection of avian viruses via illumina and nanopore sequencing

Iryna V. Goraichuk[1,2], Erica Spackman[2], David L Suarez[2]

1 National Scientific Center Institute of Experimental and Clinical Veterinary Medicine, Kharkiv, Ukraine

2 Southeast Poultry Research Laboratory, National Poultry Research Laboratory, Agriculture Research Service, U.S Department of Agriculture, Athens, GA, USA

Abundant host and bacterial sequences can obscure the detection of less prevalent viruses in un-targeted next-generation sequencing (NGS). Efficient removal of these rRNAs is vital for accurate viral detection. In this study, we conducted an evaluation of the host depletion methods aimed at enhancing the detection of avian viruses through NGS. Our investigation primarily focused on the substitution of DNase I with TURBO DNase within our established RNase H-based depletion protocols to reduce the abundance of host and bacterial reads using DNA probes designed to target chicken 18S and 28S rRNA, specific chicken mitochondrial RNA, and select 16S and 23S bacterial rRNA.

We validated the performance of the modified depletion protocol by comparing the NGS results obtained using both Illumina and Nanopore sequencing platforms. To carry out this evaluation, we utilized clinical swabs collected from SPF chickens infected with highly pathogenic avian influenza virus A/turkey/Indiana/22-003707-003/2022 (H5N1). The results from our study showcased significant reductions in host rRNA levels for both depletion protocols, with TURBO DNase demonstrating superior performance. Regardless of the depletion assay used, both led to a reduction of not only host-specific but also bacterial reads, consequently enabling a notable increase in the number of sequenced viral reads. Specifically, in samples treated with TURBO DNase, viral reads constituted 7.7 and 3.1% of the total reads, whereas only 0.4 and 0.6% of viral reads were presented in untreated samples when sequenced on Nanopore and Illumina platforms, respectively. Moreover, a substantial 7-fold and 5-fold increase in the average number of viral reads per 100k sequences was observed in samples treated with TURBO DNase compared to those treated with DNase I on both platforms, correspondently.

Our findings underscore the critical role of host and bacterial depletion for improved avian pathogen detection and advancing the field of avian disease research through NGS.

Rapid DNA target capture for high-throughput targeted germline sequencing

Technology & Application Development

POSTER 618

Eric Boyden[1], Aaron Aker[1], Jack Amaral[1], Joseph Juodvalkis[1], Arjun Patel[1], Joseph Vieira[1]

1 Molecular Loop Biosciences, Inc., Woburn, Massachusetts, United States

Targeted next-generation sequencing (NGS) is a cost-effective alternative to whole-genome and whole-exome sequencing in both research and clinical settings. However, the lengthy sample preparation protocols with standard ~16-hour hybridizations are a bottleneck for many laboratories that require faster turnaround times (TAT). Methods that employ shorter hybridizations aim to reduce TAT but often compromise data quality with severely reduced on-target rates and coverage uniformity.

Molecular Loop Biosciences' LoopCapTM target enrichment platform, which utilizes molecular inversion probes, overcomes limitations inherent to other common target capture methodologies. Unlike hybridization capture methods, LoopCap's purification-free addition-only workflow is easily scalable and automatable; and unlike amplification-based methods, LoopCap employs dense bi-stranded probe tiling that ensures uniform coverage and resistance to allele dropout. Here, we show that performance of LoopCap is robust across a wide range of probe annealing times down to 1 hour, enabling single-day sample-to-sequencing with minimal performance degradation.

Human genomic DNA (gDNA) was captured using the LoopCap DNA Target Capture Kit with a 21-gene, 59.9 kb carrier screening panel. Probe annealing times of 1, 2, 4, and 16 hours were evaluated. Libraries were sequenced using the Illumina MiniSeq with 2x151 paired-end reads and downsampled to 95k clusters per sample. Data analysis was performed according to Molecular Loop's Data Analysis User Guide v.3.

All probe annealing durations yielded mean deduplicated coverage between 164X and 178X. Coverage completeness remained consistently high across annealing times, with 96% and 95% of the target covered at ⩾50X for the 16- and 1-hour annealing times, respectively. Reads on-target ranged from 86% to 92% across all annealing times, and coverage uniformity, measured by the Fold-80 penalty, ranged from 1.51 to 1.57. PCR duplicate rates increased as probe annealing time decreased, as expected, from 7.7% to 16.3% for the 16- and 1-hour protocols, respectively.

With an industry-leading 75 minutes of hands-on time, and an accelerated workflow that produces high-quality sequencing libraries from extracted gDNA in <5 hours, LoopCap is a superior technology for a wide variety of targeted germline sequencing applications, especially those that require rapid TAT.

Reduced pipette tip consumption and sample cross talk via liquid handling optimization and automated magnet movement on robotic liquid handlers.

Robin Coope

Early in the Covid-19 pandemic, genomics laboratories internationally were challenged to support large-scale testing and related activities. In our jurisdiction, supply chain issues required the creation of an automated viral RNA extraction protocol using off-the-shelf research-grade reagents, to be deployed as a back-up process as required. Key factors we considered included the need to reduce tip consumption and for a sensitive way to detect low-level sample cross talk, below that normally detected in sequencing. Our solutions to these issues, initially created for SARS-CoV2, have subsequently been rolled out across our all our DNA and RNA extraction and library construction protocols. Here we describe strategies to minimize both cross contamination and residual tip fluid during mixing and transfers on 96 well plates. In particular, we have developed The Maglevitator, a low profile vertical actuator for magnet arrays that obviates the need to move plates on and off the magnet during bead purification, which works with both PCR plates and 2.2mL deep well blocks. Small robots like the Hamilton Nimbus generally require tip ejection to use the gripper, so avoiding the gripper allows a tip to perform more steps, assuming fluid carryover is acceptably low. This has led us to a reduction from 9 tips per sample to 3 for our Total Nucleic Acid purification protocol and from 7 tips to 3 for our core bead purification method. We also created custom methods for the Nimbus to reduce X and Y travel speeds and to eliminate unnecessary vertical head movements between aspirate and dispense steps. Taken together, these changes have eliminated well-to-well cross contamination as shown by qPCR checkerboard assays, compared to our original extraction protocol. In addition, we have adopted improvements to reduce residual alcohol after each wash step and maximize elution recovery, which have made bead clean method robust to Z-height calibration and robot head tilt issues. The latest of these improvements were implemented in early 2023 and as of this writing we have not had reported failures related to bead purifications in our clinical or main research pipelines, or observed cross contamination in sequence data since that time.

(continue to page 260)

(continued from page 259)

Whole genome sequencing (WGS) of tumor-normal paired samples is the most complete approach for comprehensive molecular diagnostics for solid tumors. Although clinically validated, wide-spread adaptation is still hampered by higher costs compared to panel-based approaches. In contrast to panel-based approaches, which involve relative expensive target enrichment reagents, the main cost component for WGS is in the sequencing itself. Therefore, sequencing technology innovations that reduce the costs of sequencing can have a very large impact on the overall test costs and hence broad clinical adoptation.

The UG100 sequencer from Ultima Genomics uses mostly natural sequencing-by-synthesis chemistry on an innovative open fluidics platform at a price tag of $1 dollar per Gigabase. With 90x tumor and 30x normal control genome coverage, reagent costs for WGS-based cancer diagnostics test is potentially reduced from a currently ˜1800-2000 dollar (Novaseq6000) to ˜400 dollar.

To assess the suitability of the UG100 platform for diagnostic purposes, we sequenced the well-studied COLO829 tumor-normal cell line pairs and performed in-depth recall and precision analyses as compared with Novaseq6000 results. Sequencing data was processed and genome-wide variation annotated with the Hartwig Medical Foundation open source data analysis pipeline for WGS-based cancer diagnostics. This pipeline was originally developed for Illumina paired-end sequencing data (2 x 150nt), but was slightly modified and optimized to handle UG100 data (single end ˜300nt).

We will report on detailed analyses for small variant (SNV, indel), copy number and structural variant detection, demonstrating high genome-wide accuracy and concordance with Illumina Novaseq6000-based sequencing.

Refining liquid biopsy: Generating more information from cell free DNA

Technology & Application Development

POSTER 620

Fabio Puddu , Tom Charlesworth, Robert Crawford, Nicholas Harding, Annelie Johansson, Ermira Leslie, Aurelie Modat, Casper K Lumby, David J Morley, Jamie Scotcher , Jean Teyssandier, Michael Wilson, Jens Füllgrabe, Walraj S Gosal, Shirong Yu, Páidí Creed

Liquid biopsy for profiling of cell free DNA (cfDNA) in blood holds huge promise to transform how we experience and manage cancer by early detection and identification of residual disease and subtype. However, a standard blood draw yields an average of only 10 ng of cfDNA, of which DNA derived from the tumour is a small minority. Therefore, we are faced with a dilemma when utilising the limited sample to obtain maximum information.

Genetic sequencing provides information on actionable somatic mutations but detection of a few loci in circulating tumour DNA (ctDNA) which constitutes a minority of the sample is challenging. 5-methylcytosine (5-mC) profiles of cancer are differential from non-cancer at many more loci and so provide a stronger signal and 5-hydroxymethylcytosine (5-hmC) profiles have been shown to be predictive of early stage cancer. Hence, genome-wide measurement of genetic and epigenetic variation at single-base resolution could enable more sensitive liquid biopsy assays for measuring genetics, 5-mC and 5-hmC cfDNA has been technically challenging. Standard genetic sequencing does not provide information on epigenetics and existing NGS methods for methylation sequencing sacrifice genetic information to measure modified cytosine and cannot discriminate between 5-mC and 5-hmC. multiomics solution is a new technology which sequences at single base resolution the complete genetic sequence integrated with modified cytosine from low nanogram amounts of cfDNA. By introducing an important high-fidelity methyltransferase during the workflow, we are able to construct an expanded information table that can be used to deconvolute A,T,C,G, 5-mC and 5-hmC simultaneously in a single read within a DNA molecule.

Using synthetic controls and reference DNA samples, we demonstrate that this method can achieve accurate measurement of 5-mC and 5-hmC as well as achieving high accuracy for SNV calling. Further we show the derivation of cfDNA fragment characteristics from sequencing reads jointly encoding genetic and epigenetic information. The simultaneous identification of multiple modalities enhances the signal available for early detection of cancer from a single, low-input sample of cfDNA.

Retaining ssDNA and nicked dsDNA during library preparation can rescue FFPE specimens for clinical protocols

Technology & Application Development

POSTER 621

Jocelyn M Lucyshyn[1], Benjamin J Kelly[1], Peter Chang[1], Shountea Stover[2], Natalie Bir[2], Diana Thomas[2], Nilsa Ramirez[2], Kathleen Schieffer[1], Catherine E Cottrell[1], Anthony R Miller[1], Elaine R Mardis[1], Richard K Wilson[1]

1 Nationwide Children's Hospital, Institute for Genomic Medicine, Columbus, Ohio, USA
2 Nationwide Children's Hospital, Biopathology Center, Columbus, Ohio, USA

Standard tissue preservation techniques of formalin fixation and paraffin embedding (FFPE) of clinical specimens result in cross-linked, nicked and degraded nucleic acids. Conventional techniques for preparing next generation sequencing (NGS) libraries require end-polishing of the termini of each template molecule to promote ligation of the platform-specific adapter. This strategy renders all molecules uniformly blunt and converts only double stranded DNA (dsDNA) into library molecules. Therefore, the single-stranded DNA (ssDNA) or nicked dsDNA molecules present in FFPE samples are lost during conventional NGS library preparation, resulting in diminished molecular diversity within the final library.

To overcome this loss, appreciable amounts of FFPE-extracted DNA are used as input into the dsDNA library protocol to rescue complexity. The input requirement for the exome protocol offered by the IGM Clinical Laboratory is 500 ng of dsDNA for FFPE processing; a considerable amount, given some tumor biopsy specimens provide insufficient extracted nucleic acid, considering the amount needed for multiple assays. Presently, more than 20% of FFPE DNA extractions do not meet the input requirement for the IGM clinical cancer protocol and are excluded from testing.

The Claret Biosciences Single-Reaction Single-stranded LibrarY (SRSLY) kit uses a single-stranded library preparation method to convert all extracted DNA (ssDNA, dsDNA, and nicked dsDNA) molecules into a dsDNA library, including degraded material from FFPE preservation. The diverse population of template DNA that can be converted into a library by the SRSLY method helps to preserve molecular complexity within the library, enriching the representation and depth of unique molecules available for exome capture, and consequently requiring significantly lower input.

The SRSLY kit was tested on IGM clinical and translational samples with FFPE DNA input of 50 ng, a 10-fold reduction compared to the current IGM NGS protocol. In our preliminary results, the on-target capture was improved by 39% and reduced duplicates by 35%, resulting in 23% fewer total reads required to achieve 250x coverage. Additionally, the improved sensitivity and detection of germline and somatic variants that was observed exhibited the potential to rescue excluded samples based on the laboratory's DNA input mass requirements.

Revolutionary microarray technology for diagnosis and management of hematological malignancies

Shahar Zirkin[1], Yael Michaeli[1], Noa Gilat[1], Topaz Kreiser Shem Tov[1], Gal Mesika Surizon[1], Kelila Freedman[1], Orit Feldstein1, Miri Neaman[2], Jasline Deek[2] and Yuval Ebenstein[2]

1 JaxBio Technologies LTD., Netanya, Israel
2 School of Chemistry, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel

Current diagnosis and monitoring of hematological malignancies rely on invasive bone marrow aspiration or lymph node biopsy. Continuous monitoring is vital for cancer patients, aiding in evaluating disease progression and detecting minimal residual disease (MRD). In addition, routine testing is crucial for pre-malignant conditions that can progress to acute malignancies, enabling the tracking of disease evolution. Therefore, a non-invasive approach is urgently needed to enable sensitive disease diagnosis and monitoring.

The SANGUINE project, an EU-funded initiative, addresses this need with a multi-center clinical trial involving 3,000 patients. The project introduces an innovative, minimally invasive blood test to detect and classify various hematological malignancies. It relies on identifying a combination of epigenetic biomarkers in DNA from peripheral blood cells and cell-free DNA using the cutting-edge technology developed by JaxBio.

Here we present a comprehensive analysis of DNA methylation and hydroxymethylation, offering a unique signature for each hematological malignancy. The methodology involves chemoenzymatic reactions to fluorescently label these epigenetic marks, followed by hybridization to a custom microarray. A machine learning algorithm analyzes the unique patterns and defines a disease-specific epigenetic biomarker panel.

After analyzing more than 150 samples, we identified a robust epigenetic set distinguishing healthy individuals from cancer cases. The analysis revealed thousands of distinct probes that classify Lymphoma with 93% sensitivity and 94% specificity, and Acute Myeloid Leukemia (AML) with 93% sensitivity and 100% specificity.

Our findings represent a significant advancement in hematological malignancy diagnosis and monitoring. This technology simplifies diagnosis by reducing the need for invasive procedures and enabling earlier detection.

Revolutionizing Assay Development: The iconPCR System's Real-Time Analysis of 96 Concurrent Variables

Chris Streck, Yann Jouvenot, Akshit Kanda, Joseph Jin, Allan Yu, Fareed Nabkel, Yassine Kabouzi, Pranav Patel

n6 Tec, Inc. , Pleasanton, CA, USA

Historically, the creation and optimization of new molecular biology assays has demanded repetitive adjustments of reagent compositions, oligo designs, and meticulous assessments of varied conditions, encompassing temperature calibrations, time durations, and cycling parameters. Such evaluations often necessitated numerous thermocycler operations to fine-tune individual steps within an instrument's protocol or thermal profile. This protracted procedure introduced several challenges: 1) The necessity for multiple, costly PCR devices to facilitate simultaneous testing, 2) Sequential trials on one apparatus, leading to extended durations and impacting stability, and 3) Difficulties due to variability between setups and PCR devices.

Leveraging the innovative iconPCR system, which offers individually controlled PCR reactions, assay developers gain unparalleled flexibility in parallel experimental design and can obtain real-time data from up to 96 reactions concurrently. The iconPCR system features 96 PCR wells, each with independent control and measurement capabilities. Compatible with standard PCR tubes and plates, this system remains agnostic to the choice of reagents or commercial kits, promoting swift design changes within a single cycle. It allows unique thermal profiles for every well, while providing temperature accuracy on par, if not superior, to existing thermocyclers. This design facilitates various PCR cycles to operate distinctly, whether in an individual well format or a user-defined zone spanning the 96-well matrix. Traditional thermocycler functions, such as PCR cycling, melt curve, and end-point analysis, are enhanced with up to 96 linearly distributed temperature gradients, variable number of cycles in each well, and auto-normalization techniques, ensuring uniform NGS library preparations. Furthermore, the inclusion of intra-run melt curve assessment aids in real-time amplicon quality checks, pinpointing issues like primer-dimer formations or other anomalies. This approach is designed to significantly elevate the efficacy of assay fine-tuning across applications, from PCR amplicons and ligation efficacy to cDNA synthesis evaluation, CRISPR screenings, Differential Scanning Fluorimetry, protein shift assays and NGS library preparations. In this context, we introduce a spectrum of analysis tools, instrument configurations, and both qualitative and quantitative metrics pivotal for refining molecular biology methodologies.

Same-slide fully automated spatial multiomics profiling of immune cells in the tumor microenvironment through integration of RNAscope™ and sequential immunofluorescence on COMET™ platform.

Technology & Application Development

POSTER 624

Arec Manoukian[1], Alice Comberlato[1W], Paula Juričić[1], Pino Bordignon[1], Alix Faillétaz[1], Anushka Dikshit[2], Emerald Doolittle[2], Rose Delvillar[2], Steve Zhou[2], Ching-Wei Chang[2], Li-Chong Wang[2]

1 Lunaphore Technologies, Tolochenaz, Switzerland
2 Advanced Cell Diagnostics, Inc., Newark, California, USA

Spatial biology methods have shed light on the complexity of the tumor microenvironment (TME) at single-cell level, revealing common and rare cell types and their intercellular interactions in a spatial context (PMID: 37535749). Hyperplex immunofluorescence allows the identification of multiple biomarkers on the same sample, enabling immune cell profiling within the TME. However, revealing in-depth TME activation status requires mapping of cytokines/chemokines, currently achievable by RNA in situ hybridization techniques on a separate serial section. The ability to spatially profile both RNA and protein markers on the same slide is needed to facilitate investigations of specific cell populations such as immune cells and their activation states (PMID: 35132261).

Here, we show a novel multiomics approach (PMID: 31700674) that integrates RNAscope™ and sequential immunofluorescence (seqIF™) to simultaneously identify RNA and protein targets on the same tissue slide. The analysis is fully automated on COMET™, an advanced tissue staining and imaging platform with precise temperature control and full workflow automation, ensuring optimal efficiency and reproducibility. Our integrated protocol combines up to three cycles of RNA detection with seqIF™, where two markers are identified in each cycle, for a final dataset including 12-plex RNA and 20-plex protein multiomics panel.

The panel is designed to identify key components of the TME such as macrophages, dendritic cells, fibroblasts, and tumor-infiltrating lymphocytes (TILs). RNA probes are selected to provide further information about the activation state of TILs by targeting key biomarkers like co-inhibitory receptors and transcription factors, together with secreted molecules such as cytokines and proteases. The simultaneous detection of RNA and proteins provides extensive insights into the TME molecular landscape, revealing co-expression patterns and relationships between RNA and proteins within individual cells. The signature of an immunosuppressive microenvironment is also studied, focusing on TIL activation states and immune checkpoint expression.

(continue to page 266)

(continued from page 265)

Our results affirm the successful implementation of the combined RNAscope™ and seqIF™ protocols on the COMET™ platform to allow same-slide spatial multiomics analysis. Preserving spatial context and intercellular relationships, this technology will help to unravel complex cellular interactions among distinct cell populations, thus opening new avenues for personalized medicine and the identification of therapeutic targets.

Scalable dual-indexed single-cell whole-genome sequencing on Ultima Genomics

Andrew Wicks, Frank Wos, James Roche, Timothy Chu, Soren Germer, Jade Carter

New York Genome Center, New York, NY

Single cell whole genome sequencing can reveal genomic intra-tissue heterogeneity, such as copy number variation (CNV) which can be observed in individual cells with low-depth sequencing and fitted to known cancer mutational profiles1,2. With potential diagnostic tools and atlas programs expanding, it is imperative to reduce single cell whole genome sequencing (scWGS) costs to make it a more attractive and accessible modality for single cell studies. We present a method leveraging Ultima Genomics (UG) sequencing technology which drives down the per cell sequencing cost of 0.1X scWGS to $0.04 USD. Our strategy combines DLP+ (direct library preparation plus)3 with customized barcode primer constructs enabling dual indexing on a unidirectional sequencing platform.

Libraries were prepared from an hTERT cell line4 using DLP+ with custom designed barcode primer constructs that are compatible with single-end reads on the Ultima platform. Cells and primers were spotted into a 5184-well Smartchip (Takara), lysed, tagmented, neutralized, and amplified in-situ. Resulting libraries were then pooled and size selected to ~400bp, to increase the probability of reading through both indexes, then converted using a UG library conversion kit and loaded onto an Ultima sequencer at 375 pM resulting in 0.1X average coverage per cell. Reads were demultiplexed via custom scripts and analyzed through an existing mondrian pipeline5 for alignment and downstream analyses. We recovered combinatorial indexes from ~75% of the final constructs, whereas Illumina custom-converted libraries yielded only ~40% recovery on the UG sequencer.

(continue to page 268)

(continued from page 267)

References:
[1] Perner J, Abbas S, Nowicki-Osuch K, et al. The mutREAD method detects mutational signatures from low quantities of cancer DNA. Nat Commun. 2020;11(1):3166. Published 2020 Jun 23. doi:10.1038/s41467-020-16974-3
[2] Funnell T, O'Flanagan CH, Williams MJ, McPherson A, Aparicio S, et al. Single-cell genomic variation induced by mutational processes in cancer. Nature. 2022 Dec;612(7938):106-115. doi: 10.1038/s41586-022-05249-0

[3] Laks E, McPherson A, Zahn H, et al. Clonal Decomposition and DNA Replication States Defined by Scaled Single-Cell Genome Sequencing. Cell. 2019;179(5):1207-1221.e22. doi:10.1016/j.cell.2019.10.02

Scaling high-throughput, multimodal single cell CRISPR screens using 10x Genomics Chromium Single Cell Gene Expression Flex

Carolyn Morrison[1], Snow Wu[1], Ryan Stott[1], Stephen Williams[1], Ian Fiddes[1], Paul Lund[1], Andrew Kohlway[1], Peter Smibert[1], Sarah Taylor[1]

1 10x Genomics, Pleasanton, CA, USA

Combining CRISPR with single cell technology can provide additional information beyond phenotypic readouts, generating paired perturbation, gene and protein expression data. However, current on-market single cell CRISPR products sometimes require upfront single guide RNA (sgRNA) design considerations to ensure compatibility. Furthermore, the throughput can limit the scale of a CRISPR screen and therefore, the number of sgRNAs profiled. 10x Genomics' new probe-based chemistry workflow, Chromium Single Cell Gene Expression Flex, overcomes these challenges by enabling custom probe spike-ins and effective scaling of the number of cells that can be profiled in a single experiment. Here we show the compatibility of the Flex workflow with a CRISPR interference (CRISPRi) screen, showcasing two different approaches.

We performed a pooled CRISPRi screen of ~7,000 sgRNAs targeting ~700 lncRNAs in a human cancer cell line stably expressing dCas9-KRAB. Following lentiviral transduction, sgRNA-containing cells were enriched by positive selection followed by flow sorting. Cells were stained with TotalSeq-C antibodies, fixed, then partitioned into 16 or 96 individual hybridization reactions with sample-specific barcoded probes. In the 16-plex configuration, cells were hybridized with a whole transcriptome probe set, in addition to a set of custom probes designed to capture the sgRNAs and their lncRNA targets. In the 96-plex configuration, hybridization reactions used a targeted gene panel in lieu of a whole transcriptome probe set. To enable separate library preparation and optimal sequencing, guide libraries were engineered with amplification handles distinct from the Flex probes. After an overnight hybridization, samples were pooled in a 16-plex or 96-plex multiplexing configuration and partitioned into GEMs (Gel beads-in-EMulsion) across 8 lanes on a single chip. The built-in multiplexing capability enabled the capture of 1 to 8 million cells per chip using the 16-plex or 96-plex configuration, respectively.

The Flex workflow enables efficient scaling and can be easily modified to capture additional sequences of interest following the custom probe spike-in guidance. Here we demonstrate how the Flex workflow can be used to scale a multimodal CRISPRi screen, by increasing the plexy and also by creating targeted gene panels to reduce sequencing costs.

Scaling single-cell sequencing: Multiplexing innovations for cost-effective high-throughput profiling

Amy Hamilton, Dominic Skinner, Nick Pervolarakis, Jerushah Thomas, Dmitry Pokholok, Sanika Khare, Aimee Beck, Ashley Woodfin, Maggie Nakamoto, Felix Schlesinger

Single-cell sequencing technology has vastly enriched our understanding of biology across various domains. However, its widespread application in high-throughput endeavors such as screens, atlases, and large cohorts has been hindered by the high costs associated with instrumentation and reagents, along with challenges related to throughput and sample diversity. Addressing these constraints, combinatorial indexing exploits the inherent characteristics of individual cells as reaction compartments, allowing sequential barcoding of samples within a plate-based workflow. This innovative approach obviates the necessity for complex and expensive instrumentation, offering a robust, cost-effective, and high-throughput protocol.

Scale Bio offers flexible avenues for sample multiplexing, accommodating a broad range of cell inputs into both the fixation workflow and the 3-level plate based single-cell RNA workflow. Scale fixation can be optimized for lower cell numbers, minimizing volumes and post-fixation losses for precious samples. The Scale low cell input fixation protocol accommodates up to 1 million cells in a 1.5mL tube, streamlining the fixation process. Complementing this capability, ScalePlex enables further multiplexing with hash oligos during the fixation protocol upstream of the Scale scRNA workflow. This versatility is particularly advantageous for scenarios involving samples collected and fixed at different times, low cell numbers, or the pooling of samples from diverse sources, streamlining sample handling without compromising data integrity and saving valuable time.

The modular scRNA plate-based workflow incorporates 96 unique barcodes in its initial step, theoretically allowing for 96 distinct fixed samples. Additionally, the Scale scRNA kit can efficiently handle cell inputs ranging from 2,500 to 10,000 cells per well, ensuring robust downstream cell recovery and sequencing metrics. Augmenting throughput capacity, Scale introduces extended throughput kits, enabling recovery of up to 500,000 cells throughout the assay, effectively encompassing the input and output ranges.

Scale Bio's comprehensive range of bespoke single-cell RNA sequencing offerings is adept at handling a wide spectrum of cell inputs, ensuring increased recovery rates, providing a variety of fixation solutions, and ultimately streamlining sample processing with the incorporation of ScalePlex. These innovative solutions empower researchers across a multitude of disciplines, enabling cost-effective and high-throughput investigations that drive the frontiers of biological knowledge.

Self-supervised deep learning enables label-free high-dimensional morphology profiling of immune cell subtypes

Thomas Vollbrecht, Vivian Lu, Kiran Saini, Senzeyu Zhang, Ryan Carelli, Jeannette Mei, Amy Wong-Thai, Kevin B. Jacobs, Stephane C. Boutet, Mayar Salek, Maddison Masaeli

Morphology as a high-dimensional cell characterization descriptor that can be integrated with other "-omics'' modalities is a powerful tool in providing a more comprehensive understanding of cell biology. Through Deepcell's innovative technology, we apply deep learning and computer vision to characterize cells in a label-free manner based on real-time interpretation of high-content morphology data.

The Deepcell REM-I platform combines microfluidics and high-speed imaging in the REM-I instrument, deep learning and computer vision in our Human Foundation Model (HFM), and data analysis and visualization in our Axon data suite. The hybrid self-supervised deep learning model produces 115 embeddings which are reproducible, quantitative, and high-dimensional descriptions that distinguish individual cells from one another. Once cell images are processed by the HFM, deep learning embeddings can be analyzed through interactive UMAPs, population analytics, and image visualizations. Using these features, we can define cell populations of interest and reproducibly use select features for subsequent experiments or sort populations into six wells for further multi-omic and functional analysis. To assess the immune cell subtypes typically present in PBMC samples, we used the REM-I platform to image patient-derived PBMC samples and purified immune cell subsets including CD4+ T cells (total, activated, and naive), CD8+ T cells (total, activated, and naive), CD14+ monocytes, CD19+ B cells, CD38+ plasma cells, CD56+ NK cells, and macrophages. High-dimensional morphology analysis showed profiled immune cell subsets in distinct locations on the associated UMAP, indicating cell types are separable by morphological features extracted by the HFM. Leiden clustering further distinguished cell groups by morphological traits, with several cell subsets overlapping with individual leiden clusters. The results suggest the observed morphology profiles may reflect specific functional states.

Self-supervised foundation model captures high-dimensional morphology data from single cell brightfield images

Kiran Saini[1], Senzeyu Zhang[1], Ryan Carelli[1], Kevin B. Jacobs[1], Amy Wong-Thai[1], Cris Luengo[1], Vivian Lu[1], Anastasia Mavropoulos[1], Andreja Jovic[1], Jeanette Mei[1], Thomas Vollbrecht[1], Stephane C. Boutet[1], Mahyar Salek[1], Maddison Masaeli[1]

1 Deepcell, Inc. Menlo Park, CA, USA

Many methods for imaging and sorting cells rely on biomarkers that may perturb cells and create selection bias. To eliminate dependence on biomarkers and enable broader assessment of cells, we developed REM-I, a platform that characterizes cells based on high-dimensional morphology. Live cells are captured with brightfield high resolution imaging and processed in real-time by self-supervised deep learning models to generate quantitative AI embeddings representative of cell morphology. One significant technical challenge in building this platform was developing an AI model to extract features from cell images of diverse human cells without prior knowledge of cell types, cell preparation, or other application-specific knowledge for an exploratory approach.

Accordingly, we developed the Human Foundation Model (HFM), which is a hybrid architecture that combines self-supervised learning (SSL) and morphometrics (computer vision) to extract 115 dimensional embeddings representing cell morphology from high-resolution REM-I cell images. SSL produces a foundation model with high generalization capacity that enables hypothesis-free sample exploration and efficient generation of application-specific models. Meanwhile, computer vision extracts features that represent measurable and interpretable concepts(e.g., cell size, shape, texture, intensity).

Here, we describe the training process for the HFM self-supervised backbone model, the discriminatory power added by supervised tasks, and validation of the reproducibility and generalization capabilities of the resulting model. We also report on the Axon data suite, a user-friendly tool designed for researchers to analyze data and create custom reusable workflows. This includes the ability to store and manage data, visualize high-dimensional data as low-dimensional projections, and train classifiers to identify and sort cell populations on the REM-I instrument. To enable compatibility with user-preferred downstream analysis pipelines, the tool provides data export options for images, plots, and embeddings. Our approach allows users of all computational skill levels to access and interpret AI-enabled morphological profiling. Potential applications of the REM-I platform include heterogeneous sample evaluation, tumor cell enrichment, cell state characterization, and multi-omic integration.

Sequencing-Based High-Throughput Imaging of Proteins and RNA in FFPE Tissue Sections

Michael Lawson, Yuji Ishitsuka, Andrew Pawlowski, Zhenmin Hong, Jake Koh, Kenneth Gouin, Nathan Ing, Richard Que, Yeoan Youn, Lucy Campbell, Sara Ding, Tony Facchini, Ryan Costello, Katelyn Nelson, Jamie Stover, Howon Lee, Sarthak Duggal, Taylor Plaziak, Kaitlin Cameron, Mimi Abdu, Daan Witters, Martin Fabani & Eli Glezer.

Singular Genomics Systems, San Diego, CA, USA

In-situ multiomics is revolutionizing research in oncology, immunology and neurobiology. However, speed and throughput are major bottlenecks for these critical studies. Here, we present data from a novel in-situ sequencing system, with sub-micron resolution and ultra-high throughput capacity, employing rapid 4-color SBS chemistry to profile RNA transcripts and proteins in FFPE tissue. The system also generates virtual H&E images, producing multi-modal spatial images of 40 cm^2 of tissue in <24 hours.

We prepared 5 um serial sections from a kidney FFPE block and used a novel tool to precisely position multiple tissue sections on a slide. Following probe binding and amplification, we used SBS readout to profile 105 transcripts and 6 proteins (via DNA-conjugated antibodies). Images were analyzed using a custom pipeline for feature detection and decoding, and cell segmentation using Cellpose. Results showed strong correlation to single-cell transcription profiles and cell type signatures defined by scRNA-Seq. Additionally, we see good agreement between single cell transcript and protein profiles.

Existing spatial biology platforms are known to struggle with degraded and challenging clinical sample types (e.g. decalcified bone marrow), due to RNA fragmentation. We successfully profiled immune cells in FFPE-embedded bone marrow biopsies, measuring 143 transcript targets and 8 proteins. We believe this represents a significant step forward for in-situ analysis of bone marrow samples, and could be valuable for studying the pathology and response to treatment in blood cancers such as AML and CML.

(continue to page 274)

Single-cell multiomic benchmarking of cryopreserved human peripheral blood mononuclear cells (cPBMCs) and cryopreserved human bone-marrow mononuclear cells (cBMMCs) across multiple locations, replicates, donors, and, users using CITE-seq.

Josh Croteau[1], Joseph Modarelli[2], Alexandra Corella[1], Priya Halvorsen[2], Anna Sheydina[1], Doug Wilson[2], Santosh Putta[1], Jon Lee[2], Kevin Taylor[1]

1 BioLegend Inc., Proteogenomics Department, San Diego, California, USA
2 Q2 Solutions, Translational Genomics, Durham, North Carolina, USA

Emerging single-cell multiomic techniques like CITE-seq (combined gene expression and cell surface protein profiling) require best practices and practical limitations fully established; especially related to clinically relevant samples. We designed a study across two organizations using matched donor cryopreserved human PBMCs and BMMCs to examine performance of CITE-seq on the 10x Genomics Chromium platform. Within cPBMCs reduced sample quality was simulated by "cold-shocking" samples during cell-thaw and compared to non-shocked samples. Within cBMMCs, identical cell-thawing and processing protocols were used to investigate reproducibility across study locations. Additionally, the impact of ADT sequencing depth was investigated on cBMMCs.

Over 140 surface proteins (ADTs) were examined in addition to >1100 RNA transcripts. Impact on 10x Cell Ranger metrics as well as staining quality aspects using TotalSeq™-C antibodies were examined. 37 canonical PBMC and 43 canonical BMMC cell-types were established by TotalSeq™-C antibody staining using traditional "flow-like" bi-variate gating. Cell-type percentages, ADT and RNA counts were extracted for each cell-type. Cell-types were also used to supervise subsequent tSNEs for high-dimensional analysis.

Intra-replicate correlations of cell-type, ADT, and RNA metrics were extremely high for all samples for both cPBMCs and cBMMCs. Cell Hashing antibody performance in cPBMCs was high quality, however, poor in cBMMCs. "Cold-shocked" cPBMCs suffered from low cell viability and reduced cell yields post-thaw. Cell Ranger ADT and RNA metrics and were notably reduced in "cold-shocked" cPBMCs as was the staining index for ADTs. Celltype, ADT, and RNA metrics had low correlation between conditions; especially within T cells. Within BMMCs ADT staining sensitivity was lower than in PBMCs, but, sufficient to establish valuable canonical celltypes including some CD45 negative "progenitor" populations. Cell-type metrics, ADT, and RNA counts between study sites were highly correlated indicating excellent technical reproducibility of CITE-seq on cBMMCs. Tripling sequencing depth improved ADT staining index on lowly expressed proteins, however, did not significantly alter ability to discriminate canonical cell-types.

(continue to page 275)

(continued from page 274)

An even deeper view of cellular immune response may be possible with in-situ sequencing of antibody genes and T-cell receptors. As a step towards this goal, we demonstrated in-situ sequencing of antibody heavy and light chains within a B cell line, resulting in accurate readout of 28 bases within the CDR3 regions.

In conclusion, we present an SBS-based approach for in-situ detection of targeted gene transcripts and proteins at an unprecedented throughput, and its application to the analysis of human kidney and bone marrow FFPE tissue samples. The capability to deeply profile tissues at high throughput could enable larger translational studies, and assembly of 3D maps of gene and protein expression at subcellar resolution.

Single-slide, in situ multiomic imaging of mRNA, protein and protein-protein interactions in the tumor-immune microenvironment of bladder cancer patients

Technology & Application Development

POSTER 633

Ge-Ah Kim[1], Sonali Deshpande[1], Li-Chong Wang1, Kate Allaire[2], Junyan He[2], Maithreyan Srinivasan[1], David Y. Oh[2], and Lawrence Fong[2],

1 Advanced Cell Diagnostics, a Bio-Techne brand, Newark, CA 94560
2 Division of Hematology/Oncology, Department of Medicine, University of California, San Francisco, CA 94143

Individual cells form interconnected communities through physical interactions of cell surface proteins that promote communication and structural integrity. The network of interactions and the underlying stoichiometry both within a tissue and with the immune repertoire is adversely impacted in many diseases, including cancer. While screening technologies to map the surface protein interactome such as SAVEX-IS (PMID: 35922511) are advancing our understanding of intercellular/intracellular protein-protein interactions (PPIs), no orthogonal technologies exist that can validate identified PPIs in the context of known mRNA/protein markers. To address this need, we have developed a novel technology that uses modifications to RNAscopeTM – a validation technology known for its high sensitivity and specificity - to visualize PPIs, proteins, and mRNA in situ on a single tissue section. Oligonucleotides that bind RNAscope signal generation complexes were conjugated to antibodies and protein targets were visualized in situ along with PPIs/mRNAs using a semi-automated workflow on a Leica BOND Rx instrument followed by whole slide scanning and analysis. The assay has the potential to interrogate up to 12 PPIs/mRNA/protein target combinations. The first iteration of the assay, optimized to interrogate one PPI and any combination of up to 11 mRNA and protein targets, assay was deployed on samples from patients with localized, muscle-invasive bladder cancer who received neoadjuvant anti-PD-L1 checkpoint inhibitor therapy prior to planned surgical cystectomy on a clinical trial (NCT02451423). To characterize tumor immune microenvironment of these patients, PD-1/PD-L1 interactions were assessed using a panel of targets to include cell phenotyping protein markers for both immune and tumor cells (CD3, CD4, CD8, CD107a and PanCK) and mRNA markers for secretory proteins (IFNG, GZMB, GZMK). Interestingly, several trial patients demonstrated evidence of PD-1/PD-L1 interactions within tumor after immunotherapy treatment, as well as increased cellular interactions between cytotoxic CD8+ T cells and CK+ tumor cells, even though some of these patients did not have global treatment-induced CD8+ infiltration into tumor based on orthogonal multiplex IHC quantitation. We will present data to illustrate the novelty of the assay as well as secondary analysis to characterize the importance of PPIs for more granular dissection of immune modulation in the tumor microenvironment.

Solution-phase single cell indexing by kinetic confinement

Lucas Edelman1, Pietro Marafini1, Dominic Smith1, Anna Lamstaes1, Raian Contreras1, Ieuan Williams1, Imogen Carruthers1, Own Ambridge1, Victoria Sanders-Brown1, Elvina Intaite1, Chun Yuan Hii1, Benjamin Hume1, Ryan Cardenas1, Uday Munagala1, Michael Stubbington1, Lucyna Zawada1, William Plumbly1, Frances Brown1

1 CS Genetics Limited, Cambridge, United Kingdom

Existing tools for single cell genomics require various complex physical frameworks for the indexing of cellular nucleic acids, including proprietary instrumentation, microfluidic or manually-generated droplet emulsions, and laborious combinatorial indexing schemes. The complexity and cost of these tools significantly constrains the use of single cell technologies for high-impact applications across basic and translational research.

Here, we introduce single cell indexing by Kinetic Confinement. This categorically-new instrument-free platform for single cell genomics leverages novel, bifunctional indexing reagents to deliver index sequences directly to single cells; and then a biophysical process known as 'Kinetic Confinement' to perform high-fidelity indexing of target molecules at the single-cell level, across thousands of cells simultaneously in one-pot, single-tube, solution-phase reactions.

We present the key technical components of Kinetic Confinement and the SimpleCell™ platform from CS Genetics, which provides kit-based instrument-free single cell analysis of gene expression and other cellular/multi-omic readouts. We show that Kinetic Confinement is compatible with a wide range of cell and specimen types, and exhibits analytic performance comparable to alternative platforms. We also demonstrate the SimpleCell™ platform in a series of representative functional studies, including the high-throughput measurement of drug targets and response mechanisms at the single cell level, multi-omic readouts combining both proteomic and transcriptomic measurement, and high-fidelity single cell immunology leveraging the versatile biochemistry of SimpleCell™ in conjunction with long read NGS.

(continue to page 278)

(continued from page 277)

We demonstrate that the exceptional versatility and unique practical features of Kinetic Confinement enable genuinely simple, fast, and flexible single cell experiments, and allow straightforward scaling of experimental designs across multiple orders of magnitude. We expect that Kinetic Confinement and the SimpleCell™ platform will significantly expand the scope, use, and human impact of single-cell analysis across fundamental and applied research, as well as within therapeutic development and ultimately applied clinical diagnostics.

Spatial Characterization of Isoform Landscapes in the Human Brain

Hope Eden[1], Luke Morina[2], Stephanie Hicks[4], Kristin Maynard[3], Keri Martinowich[3], Winston Timp[1,2]

1 Johns Hopkins University, Department of Molecular Biology and Genetics, Baltimore, MD, USA
2 Johns Hopkins University, Department of Biomedical Engineering, Baltimore, MD, USA
3 Lieber Institute for Brain Development, Department of Neuroscience, Baltimore, MD, USA
4 Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics, Baltimore, MD, USA

Both recently dubbed method of the year by Nature Methods, spatial transcriptomics and long-read sequencing stand out as platforms for genetic exploration with unrivaled resolution. Spatial transcriptomics adds tissue organization context to RNA sequencing, revealing genes with spatially distinct expression patterns, and identifying neighborhoods within tissues. Long-read RNA/cDNA sequencing captures full-length molecules enabling direct quantification of RNA isoforms. Our work aims to amplify the utility each of these methods offer by applying them in tandem. Conventional spatial transcriptomics methods coupled to short read sequencing report gene level information and do not capture splice isoforms, limiting the ability to detect differential transcript usage within tissues. As phenotypic complexity is driven by the alternative splicing of RNA into functionally distinct messages, employing methods with the sensitivity to distinguish transcript isoforms promises to fine-tune our understanding of elaborate biological systems.

We have taken advantage of the compatibility between Oxford Nanopore sequencing (ONT) and the 10X Visium spatial platform to generate spatially resolved RNA sequencing data with single-molecule resolution. The feasibility of this approach is facilitated by recent advances in ONT technology which have improved sequencing accuracy such that we see a 5% increase in per base quality scores. This enables confident demultiplexing of spatial barcodes and UMI deduplication. As proof of concept, we generated a “spatial barnyard” by applying tissues from both mouse and human brains to a Visium slide array. After processing with 10X reagents, and sequencing the resulting mixed species cDNA using ONT, we demultiplexed the reads and assessed the frequency at which a barcode known to be covered by either tissue was represented by reads that mapped to the corresponding genome. We saw > 95% species specificity from demuxed reads, validating the use of this approach for further investigations. As a result, we have profiled transcriptomes of the dorsolateral prefrontal cortex from four post-mortem tissue samples to characterize transcript isoform expression patterns across anatomically and transcriptionally distinct layers of the human brain.

Spatial genomic analysis of endothelial cells in in vitro vascular models using Element AVITI sequencing system

Technology & Application Development

POSTER 636

Dayu Teng[1], Julie Li[1], Shu Chien[1, 2]

1 Institute of Engineering in Medicine, UC San Diego

2 Shu Chien-Gene Lay Department of Bioengineering, Department of Medicine, UC San Diego

Introduction: Endothelial cells (ECs), lining the inner walls of blood vessels, are constantly subjected to fluid shear stress generated by blood flow. These ECs exhibit distinct gene expression profiles in response to varying inflammatory stimulations. In this pilot study, we employed the commercially available AVITI sequencing system and cartridge, with new flow cells and probe chemistries to profile the transcriptional regulation of 40 genes across the entire EC population, particularly under inflammatory stimulations, with subcellular compartmental resolution. This platform can be scaled up to investigate spatial genomics in ECs cultured within our 3D-printed flow chambers and exposed to precisely controlled complex flow conditions, thereby revealing the heterogeneity of EC responses to different hemodynamic conditions and their functional consequences.

Methods: ECs cultured on glass slides were subjected to inflammatory stimulations including TNF-α (50 ng/ml) and flow conditions for 24-48 hours, followed by fixation, priming, and sequencing analyses. We utilized the AVITI-based sequencing technology to profile the expression of 40 genes at the single-cell level, allowing for the spatial decoding of specific mRNAs within individual ECs.

Results: Initially, we validated the TNF-α-induced regulation of VCAM and eNOS expression in ECs using the AVITI system, confirming concordance with the qPCR results. Subsequently, we successfully quantified the gene expression profiles of the 40 genes of interest in ECs in flow models, including MMP1, ICAM1, and CXCL8. Among the mechano-transduction relevant genes, MMP1 expression demonstrated clear subcellular compartmental regulation, warranting further investigations.

Conclusions: The commercially available Element AVITI sequencing system proved capable of profiling multiplex (40-gene) spatial genomic data at subcellular resolution without requiring modifications to the instrument. The chemistry employed by AVITI offers a unique combination of benefits, including a low non-specific binding dye-labeled substrates, high-resolution imaging and analysis methods, and flexible surface chemistries. We intend to leverage this technology to explore spatial genomic regulation in ECs within our 3D-printed flow chambers, incorporating Computational Fluid Dynamics (CFD) simulation-defined flow patterns. This approach promises to yield novel insights into the hemodynamic regulation of EC biology in health and disease.

Spatial profiling of neuronal microenvironment during neurodegeneration

Technology & Application Development

POSTER 637

Yijia Sun[1], Bin Wang[1], Nicolas Fernandez[1], Hao Wang[1], Renchao Chen[1], Timothy Wiggin[1], Simran Kaushal[1], Yuan Cai[1], Yi Sun[1], Leiam Colbert[1], George Emanuel[1], Jiang He[1]

1 Vizgen Inc., Cambridge, Massachusetts, USA

Understanding the cellular composition and cell-cell interactions in Alzheimer's Disease (AD) is important for researching disease mechanisms, potentially leading to the development of effective therapeutic interventions. In this study, we used Multiplexed Error-Robust Fluorescence in situ hybridization (MERFISH) technology on the Vizgen® MERSCOPE® Platform to resolve the spatial distribution of biomarkers and cellular heterogeneity of the human AD brain by simultaneously imaging 6 protein biomarkers and hundreds of RNA species. We employed a representative 6-plex neuronal protein panel of cell type markers (GFAP, Iba-1, MBP), a vascular protein (CD31), and neuropathological markers (mOC23, AT8) in conjunction with a customized 244-plex gene panel to explore the intricate molecular and cellular signatures within affected human AD brain regions. We demonstrated that simultaneous RNA imaging and protein staining enabled by the MERSCOPE Platform revealed distinct morphological and pathological features of the disease that could not be identified by protein staining or single-cell RNA sequencing alone. For example, a reduction in GLUL and SNAP25 transcripts correlated with a large increase of Aβ deposits and TAU hyperphosphorylation, supporting reduced neuronal transmission activity and gradual loss of synaptic function in an aged AD brain sample. The spatial multi-omics profiling of both protein and RNA with MERSCOPE offers a comprehensive approach to researching protein expression profiles, cellular heterogeneity, and cell-cell interactions in AD pathology, helping to elucidate contributing factors to this complex disease. Our findings contribute to a deeper understanding of the cellular and molecular mechanisms involved in AD, potentially identifying novel targets for therapeutic intervention.

Spatial Single-Cell Whole Transcriptome Profiling of Human Tissues and Fixed Cells Using Spatial Molecular Imaging

Technology & Application Development

POSTER 638

Joseph Beechem[1], Hiromi Sato[1], Rustem Khafizov[1], Michael Rhodes[1], Tushar Rane[1], Ryan Garrison[1], John Hamann[1], Joseph Phan[1], Zachary Reitz[1], Dwayne Dunaway[1], and David Ruff[1]

1NanoString® Technologies, Seattle WA.

High-plex single-cell spatial transcriptomic technologies continue to advance, revolutionizing almost every area of life science research. With the goal of achieving the "holy grail" of spatial imaging, NanoString has developed the Human Whole Transcriptome RNA Imaging Panel at 18,936-plex, covering the complete protein-coding transcriptome in humans. Titration studies revealed that high-sensitivity whole transcriptome imaging can be obtained using 2-hybridization tiles per protein-coding gene, resulting in a hybridization panel comprising over 37,872 imaging barcodes. These barcodes are encoded at 156 bits of plexity information (4 on-cycles and 35 dark-cycles per code) and each imaging probe is Hamming Distance of 4 from each other to achieve a very low false-code detection.

Advancing single-cell imaging to the whole transcriptome level opens a single unified approach to accomplish essentially all single-cell experimentations, both imaging and non-imaging. Depending upon the sample type (e.g. fixed-cells, organoids, tissue sections, etc.), the transcripts-per-cell and genes-per-cell data of the whole transcriptome panel, often exceeds that obtained by the highest-resolution single-cell RNA-seq, with up to 15,000 transcripts per cell and over 3,500 genes per cell possible. To achieve accurate cell segmentation and transcript assignment of the whole transcriptome into individual single cells, the CosMxTM Spatial Molecular Imager has been extended to allow high-plex protein imaging (up to ~75-plex) on the same slide that was used for whole transcriptome imaging. New RNA decoding and tertiary analysis algorithms have been developed for this new data class, featuring advanced cell-typing and cell-state analysis, unprecedented 360-million-plex spatial-correlation analyses (18,936 X 18,936), full spatial biological pathway enumeration, and neighborhood analysis. The high-dimensional whole transcriptome plus multi-omic protein data is streamed directly to a cloud-based Spatial Informatics Platform. These rich multi-omic data sets, when combined with data-driven machine-learning cell-segmentation algorithms, generate the most complete view of single-cell and sub-cellular spatial biology that has ever been obtained.

Data will be presented spanning the entire range of single-cell experimentation, from cells-in-culture to dissociated-tissue samples to entire FFPE-tissue sections, with a focus on cancer biology and human brain imaging.

Streamlined target enrichment and uniform sequencing using xGen Pre-hyb Normalase technology

Katelyn Larkin, Ushati Das Chakravarty, Jennifer Bowser, Ashley Dvorak

Integrated DNA Technologies, Coralville, IA USA

In traditional hybrid capture workflows, researchers are required to spend valuable time performing quality control on each library and individually pooling varying volumes to create multi-plex pools for capture. This method is time-consuming, costly, and prone to pooling and pipetting errors. To circumvent these challenges, we have developed a novel, enzymatic library normalization technology that eliminates the need for library quantification and uses equal volume pooling prior to hybridization capture resulting in a streamlined workflow with a sample read depth CV of <10%. The Pre-Hyb Normalase (PHN) technology was thoughtfully designed for compatibility with library preparation kits that utilize indexing via PCR and kits that ligate full-length adapters, allowing for easy integration into most workflows. xGen Pre-Hyb Normalase enables a fast, scalable multiplex library normalization solution for mid-high-throughput labs that value efficiency and minimizing overall costs.

Here, libraries were generated using the xGen cfDNA & FFPE library prep kit with xGen Pre-hyb Normalase. To compare normalization technologies, the xGen technology was benchmarked against a competitor bead-based library and normalization technology. For both IDT and the competitor, libraries were pooled into 16-plex captures using equal volumes without any quantification. All libraries underwent hybridization capture using IDT's xGen Hybridization Capture Core Reagents. When comparing read count per sample, use of the xGen PHN technology resulted in a lower CV and more equal library balance than the competitor. Furthermore, the IDT PHN technology showed consistent target enrichment metrics, highlighting the quality of the workflow solution.

Unlike other normalization workflows, xGen Pre-hyb Normalase chemistry is not limited by sample-input amounts and is compatible with a wide array of xGen library preps enabling multiple NGS applications. Using only 20 ng input into library prep, PHN technology generated quality data allowing for correct variant identification in reference material. Pre-Hyb Normalase is capable of powering many NGS workflows, including xGen, and results in an easier workflow due to elimination of timely, individual QC steps, and increases cost efficiency by providing more even library coverage and less wasted sequencing reads.

Streamlined workflows with improved detection of low abundance transcripts

Technology & Application Development

POSTER 640

Laurie Kurihara, Sydney Harding, Benli Chai, Sukhinder Sandhu and Steven Henck

One challenge of RNA-Seq is the high read depth needed for comprehensive analysis and detection of low abundance transcripts. To improve performance and reduce workflow time and cost, we optimized library prep chemistries and enrichment methods. These updated workflows use an efficient ssDNA adapter attachment chemistry to directly convert first-strand cDNA into NGS libraries. This approach saves time, generates stranded libraries without requiring conventional second-strand cDNA synthesis, and maintains library complexity at low inputs since adapter titration is not required. Here, we present workflow options combining the improved library prep with two different enrichment methods: a new polyA selection bead module and hybridization capture panels for targeted sequencing. Compared to several commercially available workflows, the improved polyA RNA workflow detected significantly more low abundance transcripts, even when using a low 10 ng total RNA input- which resulted in the same number of transcripts identified from up to 50% fewer total sequencing reads. Compelling improvements were also observed for hybridization capture enrichment, when performed in place of polyA selection or ribodepletion. As demonstrated with exome capture, the input for library prep could be reduced to 1 ng total RNA, where sensitivity in detection of genes and transcripts was maintained. Benefits were also observed for workflow time by combining the 4 hour library prep with shorter hybridization which resulted in a targeted RNA-Seq workflow of only 10 hours total time without loss in performance. By using smaller, more focused hybridization capture panels, sequencing costs were also dramatically reduced. For example, a 56 gene targeted RNA fusion panel was tested using 25 ng total RNA from a FFPE fusion transcript standard, and detection of all fusions was achieved from as low as 1M read pairs, compared to 15M read pairs needed for whole transcriptome sequencing. By reducing RNA-Seq turnaround time, improving data quality and performing targeted sequencing, higher throughput sample processing can be achieved at a lower cost.

Structural and copy number variant detection from cell-free DNA using Linked Target Capture (LTC) technology

Joel Pel[1], Ehsan Haghshenas[1], Ravi Vijaya Sayta[1], Joshua Babiarz[1]

1 Natera, Inc., Austin, Tx, USA

The non-invasive analysis of cell free DNA (cfDNA) from blood is rapidly emerging as a powerful tool in the detection, monitoring and treatment of cancer and other conditions. In particular, targeted Next Generation Sequencing (NGS) has enabled the detection of many rare cfDNA biomarkers, typically SNVs and indels, in a single assay. However, many conditions such as sarcomas, harbor complex alterations including structural variants (SVs) and copy number changes (CNVs) that can be challenging to detect with existing methods. Additionally, current methods can require a priori knowledge of the alteration to enable detection.

Linked Target Capture (LTC) is a novel targeted sequencing technology, which uses physically linked capture probes and PCR primers to enrich entire cfDNA fragments by amplifying targeted molecules from library adapter sequences. 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/SVs with unknown partners, since the linked probe-primer pair can capture across an untargeted sequence. Additionally, LTC preserves start/stop locations of the original template, produces high uniformity (required for CNV detection), and can deliver high on-target rate regardless of panel size.

Here, we demonstrate the use of LTC for detecting SVs and CNVs in cfDNA applications. We evaluated commercially available reference materials, cell lines, and patient cfDNAs. LTC results were compared to reference material Certificate of Analysis, and cell lines and patient cfDNAs were evaluated with ddPCR as an orthogonal method. We report high sensitivity for both variant types, down to <1% tumor fraction, without prior knowledge of the alteration. Additionally, we report 99.9% specificity for detecting any variant across 29 genes.

Supporting new massive scale single cell research initiatives through development of automated solutions for sample preparation.

Technology & Application Development

POSTER 642

Dan Dubinsky, Cole Walsh, Tom Howd, Jim Meldrim, Sheila Dodge, Tim Desmet, Niall Lennon, Stacey Gabriel, Evan Macosko, Steve McCarroll

Broad Institute of MIT and Harvard, Genomics Platform, Cambridge, MA, USA

Single Cell and Single Nuclei sequencing is a continuously growing and powerful method of interrogating the transcriptome of individual cells. With platforms like 10x Genomics, it is possible to encapsulate thousands of nuclei in a single reaction. However, technically complex and highly manual workflows associated with tissue dissociations can limit not only the yield of available nuclei that can be targeted, but also the amount of reactions that can be batched into a single run, ultimately limiting the scale of the project that can be tackled. First, different tissue types require different methods of nuclei isolation limiting the scale (i.e. sources of samples labs that can be extracted at once). Second, the varying inputs going into the assay affect downstream PCR and sequencing depth per reaction.

We present an automation solution that has been paired with the implementation of the Illumina® NovaSeqX Plus, making it possible to receive nuclei from multiple sources and sequencing them at scale. With these automation protocols, we are able to receive, store, then batch nuclei from multiple sources at varying levels of nuclei being input into the process, and run plate based cDNA generation and generate libraries consistently at 450bp with concentrations above 2ng/uL for over 85% of samples. Then with custom automated normalization and pooling, each library gets the expected sequencing depth regardless of the quality of the nuclei being processed.

At this scale, it becomes feasible to support high throughput research projects like the Brian Initiative Cell Atlas Network (BICAN), which aims to use 3' single cell transcriptome and single cell Multiome (ATAC-Seq and transcriptome) methods from 10x Genomics across thousands of nuclei suspensions from frozen tissue collected at multiple time points and developmental stages - with each suspension potentially targeting over 20,000 nuclei per reaction. Even with complex tissues our library QC methods show appropriate gene expression levels, in addition to nucleosome structure required for epigenomic data. Our automated protocol achieves 60% recovery of complex nuclei input and sequencing of at least 20,000 reads / cell. We also maintain a high Transposition Start Site Enrichment Score (TSS) despite the varying quality of the nuclei for ATAC (Assay for Transposase Accessible Chromatin) assays. Utilizing these automated, high scale assays for 10X Genomics gene expression and ATAC assays enables consistent processing for impactful, large projects.

TET Assisted Pyridine Borane Sequencing version 1.0 - A rapid, scalable, and non-destructive sequencing approach to methylation profiling

Craig Marshall[1], Travis Sanders[1], Sam Vogel[1], Martin Ranik[1], Eduard Casas[1], Doug Wendel[1], Leo Karamanof[2], Bjarne Faurholme[2], Abre de Beer[2], Lee French[2], Ross Wadsworth[2], Kristina Giordia[1], Eric van der Walt[2], and Brian Kudlow[1]

1 Watchmaker Genomics, Applications Development, Boulder, Colorado, USA
2 Watchmaker Genomics, Applications Development, Cape Town, South Africa

Detection of 5-methylcytosine (5mC) serves as a powerful biomarker for studying and monitoring disease states, particularly in the context of cancer detection. The gold standard for accurate detection of 5mC utilizes sodium bisulfite treatment of DNA, converting unmethylated cytosine to uracil while 5mC is protected (BS-seq). BS-seq, while useful, has limitations. First, conversion of all unmethylated cytosines to thymine leads to challenges during sequencing and alignment due to the low complexity of the libraries produced. Second, BS-seq results in significant DNA loss due to the destructive nature of the chemical deamination of cytosine to uracil. Thus, the required DNA input mass for preparation of BS-seq libraries is often limiting in the clinic when precious, highly degraded samples are considered. While major improvements have been made to BS-seq, and newer technologies offer solutions to some of the aforementioned technical hurdles, no solution currently exists which offers a workflow that is nondestructive, rapidly performed, automatable, and directly converts 5mC.

Here, we present TET Assisted Picoline-Borane Sequencing version 1.0 (TAPS v1.0), demonstrating improvements to the technology originally published by Schuster-Böckler, Song, and colleagues. TAPS v1.0 converts 5mC to dihydrouracil (DHU) using a series of oxidation and reduction reactions. Ultimately, DHU is recognized as dU with dA being the complementary base incorporated by Watchmaker's proprietary polymerase yielding C-to-T conversions observed at sites of 5mC. The nondestructive conversion of 5mC in TAPS v1.0 results in significantly higher library yield and complexity than BS-seq. Using Watchmaker's proprietary polymerase evolution platform, a DNA polymerase tolerant to DHU was selected for library amplification post DHU conversion allowing for uninhibited detection of 5mC in particularly dense CpG repeats. Paired with Watchmaker's End Repair and A-tailing chemistry upstream of TAPS v1.0, clinically relevant cfDNA samples of low quality and/or quantity can be assayed in methylation detection pipelines.

To assess the performance of TAPS v1.0, human genomic DNA with control spike-ins of varying methylation statuses were used. Compared to BS-seq, TAPS v1.0 saw significantly higher final library yield with greater complexity than BS-seq. Similar results were observed in clinically relevant cfDNA samples.

The KAPA Library Preparation Portfolio allows for high quality sequencing across diverse applications on the Element AVITITM System

Sandrea Theall

The emergence of novel sequencing technologies has scientists looking to adapt current library preparation workflows to new sequencing platforms. This abstract describes the use of KAPA HyperPrep, KAPA HyperPlus, and KAPA EvoPlus Kits to prepare libraries for sequencing on the Element AVITITM System. Libraries prepared using these kits were made compatible with the Element AVITITM System using the Element AdeptTM Library Compatibility Kit and assessed for multiple applications, including microbial whole-genome sequencing, human whole-genome sequencing, and human whole-exome sequencing. Libraries prepared using KAPA library preparation kits and sequenced on the Element AVITITM System showed highly uniform coverage across multiple microbial genomes with varying GC content, and enabled variant detection with high sensitivity and precision in both human whole-genome sequencing and human whole-exome sequencing. Thus, these KAPA library preparation kits allow for the generation of high-quality libraries in a diverse set of applications to be sequenced on the Element AVITITM System.

The KAPA Total Prep Workflow Facilitates Integrated Multi-Omics Data Analysis

Technology & Application Development

POSTER 645

Jieqiong Dai[1], Mariana Kiehl[1], Amy Elias[1], Nikita Raymond Dsouza[1], Shobana Sekar[1], Alejandro Quiroz Zarate[1], Lindsey Cambria[1]

1 Roche Sequencing & Life Science, Wilmington, MA, United States

The integration of multi-omics data offers an extensive understanding and profound insights into specific biological phenomena, thus significantly advancing scientific research [1, 2]. However, acquiring multi-omics data from a single sample is often resource-intensive, requiring substantial time, material, and advanced bioinformatics expertise. Addressing these challenges, the KAPA Total Prep workflow emerges as a comprehensive solution.

The KAPA Total Prep workflow facilitates simultaneous sequencing of both DNA and RNA from a single sample within a singular tube, optimizing time and resource utilization. The workflow only takes under 1.5 days to generate sequencing-ready, target-enriched libraries from samples containing both DNA and RNA as initial input. Libraries produced using low-quality Formalin-Fixed Paraffin-Embedded (FFPE) input material exhibit comparable sequencing quality compared to those from high-quality FFPE samples. Furthermore, we present an integrated bioinformatics package tailored specifically for KAPA Total Prep, streamlining data processing and ensuring robust quality control. We implemented a pipeline that efficiently segregates RNA and DNA reads within the input data, enabling focused downstream analysis targeting distinct molecular types.

Utilizing datasets generated through the KAPA Total Prep workflow from FFPE samples, we conducted comprehensive bioinformatics analyses, substantiating the efficacy of the pipeline in accurately identifying genomic variants, fusion genes, and alternative splicing events. The KAPA Total Prep workflow provides an efficient and reliable method to harness integrated multi-omics data, offering a valuable tool for advancing biological research.

For Research Use Only. Not for use in diagnostics Procedure.

References

1. Huang S, Chaudhary K, Garmire LX. More is better: recent progress in multiomics data integration methods. Front Genet. 2017;8:84. pmid:28670325
2. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics data integration, interpretation, and its application. Bioinform Biol Insights. 2020;14:1177932219899051. pmid:32076369

The PacBio Revio: A sequencing 'Revio-lution' for the Darwin Tree of Life Project at the Wellcome Sanger Institute

Emma Betteridge[1], Craig Corton[1], Iraad F. Bronner[1], James Watts[1], Stephen Inglis[2], Carol Scott[2], Caroline Howard[3], Kerstin Howe[3], Ian Johnston1

1 Wellcome Sanger Institute, Scientific Operations
2 Wellcome Sanger Institute, Pipeline Solutions
3 Wellcome Sanger Institute, Darwin Tree of Life Programme

The Wellcome Sanger Institute is a renowned genomics research institute, known for its pioneering work in DNA sequencing and large-scale genomics projects. One of its flagship programmes is the Darwin Tree of Life (DToL) Programme, which aims to generate platinum standard reference genomes of all 60,000 known species in the United Kingdom. This initiative aims to advance our understanding of life on Earth by unravelling the genetic blueprints of a wide range of organisms.

To generate the sequence data required to assemble all these de-novo genomes the Sanger Institute primarily utilises PacBio's long-read sequencing technology. With the introduction of the PacBio Revio, the institute procured multiple sequencers, to reduce cost, enhance capacity, and stay at the forefront of sequencing innovation. Building on our prior experience with earlier PacBio sequencing instruments, we designed several validation experiments, to ensure the Revio met its specifications, to understand its advantages, resilience, and constraints when subjected to the exceptionally varied DToL samples.

In addition to the validation, we incorporated these novel state-of-the-art sequencers into Sanger's production ecosystem, incorporating LIMS tracking, automated data curation and storage. Here, we will present our collaborative work on the insights, discoveries, and evaluations of our 'Reviolutonary' new Revio systems.

The Respiratory Virus and Microbiome Initiative: Expanding genomic surveillance capabilities at the Wellcome Sanger Institute

Diana Rajan[1], Emma Betteridge[1], Sarah Frances Field[1], Katie Bellis[1], Josef Wagner[1], Duncan Ng[1], Andrea Frick-Kretschmer[1], Matthew Forbes[1], Carol Scott[1], Sara Stott[1], Fernanda Novaes[1], Adrianne Lignes[1], Ian Johnston[1], David Bonsall[2], Ewan M. Harrison[1]

1 Wellcome Sanger Institute, Cambridge, UK
2 Wellcome Centre for Human Genetics, Oxford, UK

The Respiratory Virus and Microbiome Initiative (RVI) aims to establish new tools to sequence clinically important respiratory pathogens, with the long-term goal of developing routine genomic surveillance capabilities that are also capable of discovery of novel pathogens.

One arm of the RVI is focused on developing targeted sequencing approaches for the analysis of multiple common respiratory viruses (e.g. SARS-CoV-2, Influenza, RSV), in a single assay.

Here we describe our experience of three viral bait capture metagenomics workflows, utilising a combination of library construction methods and viral bait panels. These include an Illumina workflow (Respiratory Virus Enrichment Kit), a method developed by the Bonsall group (Oxford, UK), and an "in-house" method - an amalgamation of several existing sequencing pipelines established at Sanger. The bait panels tested were the Respiratory Virus Research panel (TWIST) and the Respiratory Virus Oligo panel (Illumina).

Key considerations in this development phase included:
assay tolerance to sample input quality (degradation) and quantity (low viral titre)
assay sensitivity, versus our high throughput COVID-19 amplicon sequencing pipeline
data quality - ability to call lineages, assemble viral genomes
protocol accessibility & ease of adoption by other users globally
ease of automation
amenability to cost reduction

The next phase of this project involves implementing a scalable workflow that will allow us to begin sequencing clinical samples. This exciting stage will facilitate insights into the transmission and evolution of known respiratory viruses, bringing us closer to realising one of the central ambitions of RVI - to have a robust system in place that enables us to monitor respiratory viruses and to detect new viral threats to help prevent future pandemics.

The use of deep learning algorithm development to provide automated and unbiased region selection for spatial transcriptomics data collection

James Mansfield[5], Habib Sadeghirad[1], Ning Liu[2], James Monkman[1], Chin Wee Tan[2], Caroline Cooper[3], Jeni Caldera[5], Sarah E. Church[4], Fabian Schneider[5], Ken O'Byrne[3], Melissa Davis[2], Brett Hughes[6], Arutha Kulasinghe[1]

1 The University of Queensland, Brisbane, Australia
2 The Walter and Eliza Hall Institute (WEHI), Melbourne, Australia
3 The Princess Alexandra Hospital, Brisbane, Australia
4 Nanostring Technologies, Seattle, WA, USA
5 Visiopharm, Horsholm, Denmark
6 The Royal Brisbane and Women's Hospital, Brisbane, Australia

Spatial transcriptomics (ST) enables the simultaneous visualization and analysis of gene expression patterns within tissues, providing critical insights into the spatial organization of cells and their functions, which is essential for understanding complex biological processes and diseases. One of the most time-consuming and difficult aspects of ST is the selection of the regions (ROIs) from which to collect data. We employed a deep learning / A.I. paint-to-train methodology to select regions on either serial-section H&E images or immunofluorescence images to aid in ST data collection. Unbiased and automated methods for ROI selection will improve throughput, reduce pathologist's time, and reduce bias in selection.

This study profiled pre-treatment tissue samples collected from R/M HNSCC patients (n=30) treated with Pembrolizumab or Nivolumab. Targeted spatial proteomics profiling was performed using GeoMx® Digital Spatial Profiler (Nanostring). To develop an informed region-of-interest (ROI) selection strategy for the GeoMx, merged serial-section H&E imagery and 4-channel GeoMx immunofluorescence (IF) images were analyzed using A.I.-based deep learning approaches in Oncotopix® Discovery (Visiopharm) to find region of high, mid, and low tumor content, which were each subdivided into epithelial/stromal regions (areas of illumination, AOI). In each region, 80 proteins were analyzed using the GeoMx. We performed differential expression (DE) analysis of the proteins localized in the tumor and stromal compartments against response to therapy parameters of complete, partial, stable and progressive disease, with many up- and down-regulated proteins correlating to responder population. Analyzed GeoMx IF images provided per-ROI and per-AOI cell counts (total and phenotyped based on 3 markers), and these data were used to aid in area and cell count normalization/deconvolution methods.

(continue to page 293)

(continued from page 292)

We found that an automatable, deep-learning-based ROI selection strategy based on H&E or IF images aided in defining key features in the tumor microenvironment to improve the workflow for ST data collection, that per-ROI pathology analysis of images improved comparative analysis across samples and that per-region cell counts aided data normalization/deconvolution methods. Moreover, we found that immune checkpoints and proteins involved in cell death signaling play important roles in the immune responsive tumor microenvironment of HNSCC based on response to immunotherapy.

Thinking beyond the benchtop instrument: use of integrated automation to enhance genomics workflows

Daniel R Leonce[1], Ashley S Fowler[1], Vladimir Khon[2], Alice Tome-Fernandez[2], Vicky Dearing[1], Russell Green[2], Emma Norris[2], **Jerome Nicod[1]**

1 Advanced Sequencing Facility, The Francis Crick Institute, London, UK
2 Automata Technologies Ltd, London, UK

Automation instruments are now a common feature of genomics labs and have become an integral component of modern research endeavours. Their use substantially increases experimental throughput, reproducibility, and resources optimisation. However, the scalability of automation is limited by the current approach of using benchtop instruments as standalone units. Collaborating with Automata (www.automata.tech), we have leveraged their LINQ platform to establish full walk-away genomics workflows by integrating our existing instrumentation. This completely removes manual intervention, saving scientists time, reducing error risks, and enabling equipment utilisation beyond regular hours.

The Advanced Sequencing Facility (ASF) at the Francis Crick Institute offers extensive genomics support, spanning from standard high throughput sequencing to pioneering technology applications. While our scientists invest considerable time in developmental projects, we must still provide swift services. By fully automating selected services, we have diminished the impact of routine workflows on our operations, allowing us to dedicate more time to intellectually demanding tasks. The system currently integrates a range of equipment, including a Hamilton STAR liquid handling platform, a thermo cycler, a centrifuge, and a plate reader. Its adaptable design permits future expansion for increased capacity.

The LINQ platform's manufacturer-agnostic nature facilitated seamless integration with our existing lab equipment. Additionally, its open architecture allows instruments to function independently or as part of integrated workflows. LINQ workbenches are compact, making them suitable for standard lab spaces. The LINQ software orchestrates custom workflows and oversees all platform processes. Presently, we have fully automated three workflows: CRISPR validation through amplicon sequencing, RNA extraction from FFPE samples, and SARS-CoV-2 sequencing using the ARTIC protocol. These are complete walk-away workflows, necessitating only 30-45 minutes of setup time and requiring no further manual intervention until completion 3 to 8 hours later.

(continue to page 295)

(continued from page 294)

This automation initiative has significantly increased our facility's capacity. It has liberated our scientists to focus on more demanding projects and will enable us to run extended workflows overnight, ensuring instrument availability during the day. Our next collaboration with Automata involves integrating a plate sealer and de-sealer, alongside an SPT Dragonfly, to further enhance system throughput and add new capacities.

Twist pan-cancer reference standards V2: Enhanced precision and reduced errors in ctDNA analysis

Technology & Application Development

POSTER 650

Lydia Bonar[1], Patrick Cherry[1], Shawn Gorda[3], Michael Bocek[2], Derek Murphy[2], Esteban Toro[2]

1 Twist Bioscience, Research & Development, South San Francisco, California, USA
2 Twist Bioscience, Research & Development Bioinformatics, South San Francisco, California, USA
3 Twist Bioscience, Office of Product Management, South San Francisco, California, USA

Liquid biopsies harness cell-free DNA (cfDNA) in the bloodstream which can detect cancer in its early stages. However, the precision required to discern between malignant and benign genetic markers, and the sensitivity needed for disease monitoring, make their development challenging. The formulation and utilization of specific reference materials are crucial for the successful development of these assays.

Twist Bioscience initially introduced the Twist Pan-Cancer Reference Standards*, serving as a cfDNA analog reference. This material featured a broad range of ctDNA variants, closely emulating the size and distribution of cfDNA fragments. It included synthetically-printed variants like single nucleotide variants (SNVs), insertion-deletions (INDELs), and structural variants (SVs). The reference standard includes over 400 variant sites across 84 genes, including literature-curated, clinically-relevant variant sites. Despite its comprehensive features, the first version had its drawbacks. The method used to derive the wild type (WT) background from donor derived cfDNA led to the introduction of artifacts, which limited the ability to test for low VAFs (<0.1%) at certain sites.

We have developed an enhanced version of the Twist Pan-Cancer Reference Standards, which offers a stable background genotype and a reduced background error rate. Evaluating this new reference standard with the Twist Standard V2 target enrichment system and custom capture panels targeting clinically-relevant variants, there is a clear separation between measured VAF of the lowest variant-positive standard and WT background. In addition, a side-by-side analysis showed that the error rate of this standard (V2) is markedly lower than both its predecessor (V1) and other market competitors. Importantly, its accuracy aligns closely with that of native cfDNA, and we demonstrate that it is an appropriate substrate for 'limit of blank' studies of a given variant detection assay. The updated reference materials were created at multiple variant allele frequency (VAF) dilutions, and characterized using droplet digital PCR (ddPCR) and Next-Generation Sequencing (NGS).

In summary, the Twist Pan Cancer Reference Standards V2 emerges as a valuable asset for those using NGS-based assays, paving the way for more effective liquid biopsies in clinical research.
*For Research Use Only.

Ultra high-plex single-cell spatial multi-omic imaging of human brain sections

Ashley Heck[1], Kimberly Young[1], Alyssa Rosenbloom[1], Aster Wardhani[1], Max Walter[1], Giang Ong[1], Rustem Khafizov[1], Edward Zhao[1], Sangsoon Woo[1], Lidan Wu[1], Mithra Korukonda[1], Zachary Lewis[1], Tien Phan-Everson[1], Shilah A. Bonnett[1], Claire Williams[1], Patrick Danaher[1], Rachel Liu[1], Jose Ceja Navarro[1], Christine Kang[1], Melanie Moon[1], Jim Forrester[1], Trieu My Van[1], Jason Reeves[1], Gary Geiss[1], David W. Ruff[1], Joseph M. Beechem[1]

1 NanoString® Technologies, Seattle, WA

High-plex spatial imaging of the complex cell shapes in the human (and mouse) brain biology is an unsolved problem. Brain cells often have intricate projections that can extend far from their cell body and escape segmentation by existing algorithms. As a result, roughly 50% of transcripts are left unassigned to cells when brain imaging. Also, both neurons and glial cells have localized RNA translation in their projections, enabling neural communications to stay highly dynamic. Thus, improving brain cell segmentation is essential for a better understanding of neural cell functions.

The Spatial Molecular Imager (SMI) solves this problem by enabling high-plex detection of both proteins (to image the neural projections) and RNAs on the same slide to improve brain cell segmentation and assignment of transcripts to single cells. In this multi-omic system, 68 proteins and over 6,000 transcripts covered by the Human 6K Discovery Panel are targeted on the exact same FFPE human brain section by leveraging cyclic in situ hybridization chemistry. The protein targets focus on the diverse cell types found in the brain and neurodegenerative disease pathology, whereas the RNA targets include >4,900 neuroscience-related genes focused on >80 pathways, robust cell typing, and key ligand-receptor interactions. In the multi-omic workflow: Step-1. High-plex SMI protein panels are imaged, Step-2. High-plex RNA targets on the same slide are exposed then hybridized, and Step-3. Same-slide RNAs are imaged.

We showcase this new method by a detailed examination of human brain section arrays, allowing many areas of the brain to be examined simultaneously. High-plex (68-plex) proteins, including key morphology markers GFAP, IBA1, and MAP2, are imaged simultaneously with 6,000-plex RNAs. Results highlight unparalleled segmentation of extended cell processes for neurons, microglia, and astrocytes, including over 200-micrometer-long segmented neuronal projections, the longest reported segmented cell ever accomplished using high-plex imaging. Through this CosMxTM SMI multi-omic approach, we "rescue" many of the 50% unassigned orphan RNA transcripts and quantitate the sub-cellular spatially localized expression in the human brain, an area of gene expression and biological activity that has never been quantitated at high-plex previously.

Ultra-high throughput sequencing for synthetic biology discovery

Technology & Application Development

POSTER 652

Jack T. Leonard[1], Rebecca Feeley[1], Stella Huang[1], Jenna Couture[1], Vivian Tam[1], John Palys[1]

1 seqWell, Inc., R&D, Beverly, MA, USA

Industrial-scale protein engineering and gene editing techniques have simultaneously necessitated and fueled breakthroughs in highly multiplexed molecular biology. The growth rate of NGS data output has far-outpaced development of the technologies necessary to power the Design-Build-Test-Learn (DBTL) cycle.

We show that starting from bacterial colonies, several thousand plasmids can be sequenced on a single Illumina NextSeq 2000 run and bioinformatic analysis completed by the next day.

We leverage a scalable, modular pipeline that includes automated colony picking, liquid handling, DNA sequencing and bioinformatic analysis. Rolling Circle Amplification (RCA) or direct PCR from colonies is substituted for conventional bacterial culturing and plasmid purification. Integration of an auto-normalizing, one-step library preparation technology provides 6,144 indexes in a convenient 384-well, assay-ready configuration (384 wells x 16 plates). Our benchmarking data against other workflows demonstrate that this automated pipeline shortens the typical synthetic biology DBTL cycle from weeks down to hours.

Unlocking the Potential of Dried Blood Spot Samples for Clinical Whole Genome Sequencing

Technology & Application Development

POSTER 653

Michelle Cipicchio[1], Evan McDaid[1], Dania Zaiter[1], Sophie Low[1], Brendan Blumenstiel[1]

1 Broad Clinical Labs, Broad Institute of MIT and Harvard, Cambridge MA, USA

Dried blood spot (DBS) samples present advantages for sample collection, storage and transport as compared to other sources of genomic DNA such as whole blood or saliva. This collection method is less invasive than venous blood draw, and simpler to obtain than saliva for some individuals such as pediatric patients. The collected DNA is stable long term at room temperature, enabling sample collection and shipment from areas with limited resources. Additionally, DBS samples are already routinely collected and banked for newborn screening for millions of neonates each year in the United States, representing an enormous potential resource for research.

Previously DBS samples have posed challenges when used for next generation sequencing workflows. One major obstacle can be obtaining enough high quality genomic DNA, preventing the use of PCR-free NGS methods for whole genome sequencing. Therefore most prior studies have produced whole exome or other PCR-based sequencing, where bias can affect variant calling. Through optimization of a highly scalable fully automated DNA extraction method, we have found that the majority of DBS samples can yield DNA of adequate input and quality for clinical PCR-free whole genome sequencing with at least 30X coverage. Our lab has a proven system for automated blood card punching and sample tracking that previously processed over 40,000 samples for malaria research. We leveraged this highly scalable system and built liquid handling automation for subsequent DNA purification on a Hamilton Starlet using a high throughput blood extraction kit.

Here we present results from extraction optimization experiments on contrived DBS samples, followed by results from a pilot set of over 100 real-world DBS samples collected in the field. We found that over 92% of real world samples produced more than 250 ng of DNA, meeting the standard minimum input for PCR-free WGS here at Broad Clinical Labs. We assessed DNA quality, quantity and purity, library yield, and sequencing quality metrics. We conclude that this workflow can enable clinical WGS at scale for DBS samples, opening the door for both rapid turnaround clinical tests and large scale research studies from this sample type.

Visium HD enables whole-transcriptome spatial profiling at single cell scale resolution in formalin-fixed, paraffin embedded (FFPE) tissues

Technology & Application Development

POSTER 654

Monica Nagendran[1], Jerald Sapida[1], Govinda Kamath[1], Joey Arthur[1], David Patterson[1], Augusto Tentori[1]

1 10X Genomics, Pleasanton, CA.

Technologies that spatially profile the whole transcriptome in tissues are enabling advancements in basic and clinical research. Increased spatial resolution to profile at single cell scale, higher tissue coverage, and accuracy of transcript localization are required to fully realize the potential of spatial technologies. The High Definition Visium Spatial Gene Expression assay (Visium HD) addresses these requirements and allows whole transcriptome profiling of FFPE tissues at single cell scale resolution. The Visium HD array consists of ~12 million 2 x 2 µm spatially-barcoded areas, a ~25 fold reduction in feature size compared to standard Visium arrays. The HD array has no gaps and data is binned to 8 x 8 µm for analysis. The small feature size and high tissue coverage, due to a lack of space between barcoded features in the HD array, ensures uninterrupted profiling of finer anatomical features. Visium HD assay runs exclusively on CytAssist, allowing users to prepare, stain (H&E or Immunofluorescence), and image their tissue of interest on standard glass slides using standard histological workflows. This simplifies sample handling and provides flexibility to prescreen and select tissue sections upstream of spatial profiling. Spatial resolution is dictated by a combination of feature size and extent of transcript mislocalization. We designed our HD array, workflow and consumables to ensure accurate transcript localization. By minimizing transcript mislocalization, we are able to utilize the improved feature size and ensure single cell scale resolution.

We applied Visium HD to human breast cancer and mouse small intestine FFPE tissues. Based on the localization of marker genes and unsupervised gene expression clustering data, several important cell types were identified at their expected spatial localizations. The high resolution of this technology allowed us to segregate EPCAM+ breast epithelial cells into basal (KRT14+) and luminal (KRT18+) cells. In the mouse small intestine, we were able to identify rare tuft cells. The combination of single cell scale resolution, higher tissue coverage and accuracy of transcript localization, makes Visium HD an effective tool to deliver single cell scale, whole transcriptome discovery power to a range of biological applications in basic and clinical research.

Whole-Genome Sequencing Improves Variant Detection Yield in FFPE Samples Over Whole Exome Sequencing

Technology & Application Development

POSTER 655

Mark Wang[1], Vitor Onuchic[1], Shannon Kaplan[1], Duke Tran[1], Theo Heyns[1], Li Liu[1], James Han[1]

1 Illumina, Inc, Research & Development, San Diego, CA, United States of America

Tumor profiling from FFPE specimens using next generation sequencing has been widely adopted to guide therapy decisions and support oncology research. To date, high sequencing costs have motivated labs to create small panels (approximately 15-500 genes) focused on tumor driver genes, variants that predict therapy response, and commonly reoccurring variants. Lately, research labs have started to implement Whole-Exome Sequencing (WES). Innovations in sequencing technologies continue to reduce sequencing cost and increase data yield from sequencing runs. For instance, an Illumina NovaSeq™ X run now produces 16 Tbp per run at a cost of ~$200 per 30X whole-genome with the 25B flow cell. These technology improvements are making whole-genome sequencing (WGS) a viable replacement for small panel and whole-exome based FFPE sequencing. In addition to providing a broader view of the genome for tumor profiling applications, FFPE WGS enables new applications in oncology research such as tumor informed WGS Molecular Residual Disease (MRD) testing and WGS neoantigen profiling.

This study compared WES and WGS by profiling three matched pairs of tumor and normal FFPE samples with high coverages (⩾ 275X WGS and ⩾ 300X WES) to understand the impact WGS on variant detection sensitivity. Sequencing coverage needed to achieve similar detection sensitivity of low allele frequency variants was investigated. WES reads were down sampled to 300X and WGS reads were down sampled to various levels of coverage between 50X and 275X, maintaining the matched normal samples at 60X. Between 90-97% of variants detected in WES with greater than 5% allele frequency could be detected in 225X WGS. Further, ~33% more variants were detected in coding sequence regions (CDS) with 225X WGS than 300X WES. Higher coverage uniformity in WGS is responsible for the lower coverage requirement of WGS: >99% of CDS bases were covered above 50X with WGS at 225X, while WES at 300X had 5-30% of CDS bases with <50X coverage. Reduced bias in WGS can be attributed to fewer PCR cycles and variability introduced from capture probes in exome workflows. These data suggest FFPE WGS is now a viable cost-effective workflow for profiling whole tumor genomes.

Sensitive analysis of clinically actionable genomic regions with an integrated target enrichment and sequencing workflow

Technology & Application Development

POSTER 656

Shawn Levy[1], Bryan Lajoie[1], Junhua Zhao[1], Adeline Mah[1], Laurie Edoli[1], Conner Thompson[1], Sinan Arslan[1], and Mike Previte[1]

1 Element Biosciences, San Diego, CA, USA

Over the last 15 years, hybridization-based approaches to enrich genomic regions of interest have enabled efficient and sensitive sequencing methods. From sequencing the exome for disease variant discovery to targeting small numbers of genes at high depth in somatic applications, many options exist for translational and clinical analysis. As sequencing has become less expensive, unbiased approaches like genome-wide sequencing have become common in precision health applications. Although genomes have high value and high impact, there are many applications that benefit from targeted enrichment approaches to increase sensitivity in challenging regions of the genome or to improve sensitivity to low-frequency variants, particularly in somatic applications. A significant challenge to targeted workflows is the additional technical burden and time of the hybridization workflow. These steps add several hours to a day to experimental workflows and can be challenging to automate. To address these challenges and create a highly efficient and dynamic workflow, we have created an integrated, on-flowcell target enrichment capability integrated into the standard sequencing workflow on the Element AVITI platform. This workflow leverages unique surface chemistry and does not add time or steps to the sequencing process. It also does not require any library preparation modifications. A single library preparation workflow can be used with a wide variety of target enrichment experiments. We demonstrate efficient target enrichment over regions of interest in oncology and exome. With tunable enrichment levels from 300-fold to over 1,200-fold, our platform can achieve less than 1% allele sensitivity in target regions while also providing either high specificity or uniform background coverage over the sample library at 4x to 30x levels, depending on sequencing depth. Combined, the target enrichment capabilities allow for focused and efficient rare variant detection while uniform background coverage enables variant discovery across a broad range of variant classes, including CNV. Since the enrichment methods are library preparation independent, libraries that allow more complex analyses such as structural variant detection from Hi-C libraries can be used in the protocol. These capabilities allow for advancements using this technology in numerous applications such as cell-free DNA, infectious disease, somatic variant detection and residual disease monitoring.

Characterizing and addressing error modes to improve sequencing accuracy

Technology & Application Development

POSTER 657

Semyon Kruglyak[1], Andrew Altomare[1], Mark Ambroso[1], Vivian Dien[1], Bryan Lajoie[1], Kelly N. Wiseman[1] and Matthew Kellinger[1]

1 Element Biosciences, San Diego, CA, USA

The prevailing type of error encountered in most short read sequencing methods is base substitution errors, also known as mismatches. These errors arise from various stages in the sequencing process, including library preparation, sequencing itself, and analysis. Examples of library preparation errors include deamination damage (leading to C to T mismatches) and the formation of chimeric reads (leading to split read alignments, anomalous insert lengths, or alignment induced mismatch errors). Such errors are typically assigned high base quality scores, exacerbating their impact. Mismatches can also occur in sequencing cycles due to low signal to noise or the accumulation of phasing and prephasing. In such cases, various predictors of quality are used to mark the mismatches as low quality so that they are ignored in downstream analysis. Finally, accurately sequenced bases can be incorrectly labeled as mismatches if there are differences between the genetic material being sequenced and the published reference or if reads are improperly aligned.

In this work, we quantify and attribute mismatch errors to their likely cause (library preparation, sequencing, or analysis), and then propose methods to reduce each error type. To enable accurate stratification, we begin with a PCR-free E. coli library, where the reference is well known, and no variants are expected. We then modify steps in the library preparation process to minimize deamination and detect and filter chimeric reads. For sequencing, we develop custom recipes that increase signal to noise and identify novel enzymes that reduce phasing and prephasing. We identify discrepancies with the reference by sequencing the same sample with multiple technologies. We next combine all of the improvements into a sequencing run and show a greater than 3-fold reduction in errors and demonstrate via Q score recalibration that the majority of the data exceeds Q50 (1 error in 100,000 bp). Furthermore, most of the remaining sequencing errors show low base quality, limiting their impact. Finally, we generate a long insert library of the well characterized HG001 human genome and use the improved sequencing recipe to achieve mismatch rates similar to the frequency of human variation and a 25% reduction in variant calling errors.

(continue to page 304)

(continued from page 303)

This study tested a fully-automated, self-contained system for NGS library preparation, target enrichment, pooling, normalization, and quantification. Performance was evaluated across multiple independent laboratories and several instruments, and against manual methods, for preparation of whole-exome libraries using various conditions and input types (NA12878, blood-derived hgDNA, and cell line hgDNA). WES libraries were prepared on a NGS automation technology using NA12878 hgDNA and KAPA HyperExome Probes* (50 and 100 ng inputs, n=24; 5 automated runs) in two different laboratories (A and B), resulting in singleplex and 8-plex library pools. Automated and manual library prep performance of the KAPA HyperCap v3.0 workflow* using KAPA HyperExome Probes (100 ng NA12878 hgDNA; n=24; singleplex) were compared. At sites A and B, additional samples were prepared using human whole blood and laboratory cell lines (100 ng, n=24; 2 automated runs). Variant detection by the automated workflow was evaluated with NA12878 gold standard with 24901 high-confidence SNPs in the KAPA HyperExome target region. Inter-site and inter-instrument reproducibility for site A and B was evaluated against two other sites (seven runs, five instruments). Key sequencing metrics were considered to assess repeatability of results for individual instruments at sites A and B (3 automated baseline runs, 100 ng input). The results were similar for each run: percent reads on-target >86.3% (SD<0.4 across all instruments), mean depth of coverage >54X (SD<1.1), and duplication rates of <5.9% (SD<0.6). Metrics for other automated runs were similar (lower input [50 ng], multiplexed samples [8-plex], and other sample types [blood and cell-line DNA]), as well as for manual preparation. When reproducibility was assessed across six instrument sites, greater than 85% reads on-target was observed across all instruments. This fully automated workflow demonstrated robust, consistent inter-site performance against a manual workflow for WES using the same exome probe panel. Thus, this method can reduce the hands-on time and user-to-user variability common in manual methods resulting in greater laboratory efficiency and increased confidence in the results.

nanoString®

“PLEX YES”

High Plex, Rich Biology.
Elevate your spatial insights.

Explore unparalleled biological insights with our high-plex spatial biology platforms. From groundbreaking biomarker discoveries to deciphering pathways and cell states, unlock the depths of biological understanding. **Experience the Power of Plex in Curacao 1.**

AGBT2024™

PRECISION HEALTH

SEPTEMBER 4-6, 2024

THE GAYLORD ROCKIES, DENVER, COLORADO

SAVE THE DATE

LEARN MORE BY SCANNING THE QR CODE,
OR VISIT WWW.AGBT.ORG

Index

	Last Name	First Name	Organization / Company	Page(s)
G	Gilissen	Christian	Radboudumc Research Institute for Medical Innovation	97,149
	Giorda	Kristina	Watchmaker Genomics, Applications Development	134
	Gong	Dennis	Massachusetts Institute of Technology	231
	Gonzalez-Escalona	Narjol	U.S. Food and Drug Administration	114
	Goraichuk	Iryna	National Scientific Center Institute of Experimental and Clinical Veterinary Medicine	257
	Gosal	Walraj	biomodal Lt	58,177,242,261
	Gregory	Simon	Duke Molecular Physiology Institute at Duke University	27,31,36
	Grim	Christopher	Center for Food Safety and Applied Nutrition	181
H	Hamilton	Amy	Scale Biosciences	270
	Han	James	Illumina Inc	55,69,72
	Hartwig	Anna	Exai Bio	53
	Hastie	Alex	Bionano Genomics	205
	He	Mengxiao	KTH Royal Institute of Technology	34
	Heider	Margaret	New England Biolabs	49,252,256
	Henley	Robert	10x Genomics	222
	Hernandez Gonzalez	Maria	The Steve and Cindy Rasmussen Institute for Genomic Medicine	126,179
	Hoang	Margaret	NanoString® Technologies	129,209
	Hocke	Emily	Duke Molecular Physiology Institute	27,31,36
	Hogan	Chad	Icahn School of Medicine	116
	Huang	Bo Shun	Dxome Inc.	128
	Hunt	Toby	European Bioinformatics Institute	78
I				
J	Jain	Vaibhav	Duke Molecular Physiology Institute at Duke University	27,31,36
	Jain	Komal	Frederick National Laboratory for Cancer Research	76
	Janzen	Evan	New England Biolabs Inc.	43,101

	Last Name	First Name	Organization / Company	Page(s)
K	Kanai	Akinori	The University of Tokyo	56
	Kaper	Fiona	Illumina Inc.	208,234
	Karim	Noreen	Integrated DNA Technologies, Inc.	168
	Kasinskas	Rachel	Roche Diagnostics	235
	Kelly	Benjamin	Nationwide Children's Hospital	74,126,203,262
	Kennedy	Katherine	BioSkryb Genomics	40
	Kim	Dae	Dxome Inc.	128,241
	Kim	Ge-Ah	Advanced Cell Diagnostics	276
	Kong	Mei Suen	Genome Institute of Singapore, Agency for Science	29
	Kronenberg	Zev	PacBio	167
	Kruglyak	Semyon	Element Biosciences	303
	Kucher	Natalie	Johns Hopkins University	88
	Kulasinghe	Arutha	The University of Queensland	42,122,292
	Kurihara	Laurie	Integrated DNA Technologies	202,284
L	Laliberte	Julie	Singleron Biotechnologies GmbH	157
	Langhorst	Bradley	New England Biolabs	43,49,54,71 113,252,256
	Larkin	Katelyn	Integrated DNA Technologies	202,215,283
	Lawson	Michael	Singular Genomics Systems	273
	Leonard	Jack	seqWell, Inc.	131,298
	Levine	Max	Isabl Inc.	47
	Levy	Shawn	Element Biosciences	254,302
	Li	Qiuhui	Johns Hopkins University	148
	Liachko	Ivan	Phase Genomics, Inc.	45,111
	Lithwick Yanai	Gila	Ultima Genomics	232
	Liu	George	Beltsville Agricultural Research Center	112
	Liu	Pingfang	New England Biolabs	49,252,256
	Lu	Vivian	Deepcell, Inc	182,230,246,272
	Lucyshyn	Jocelyn	Institute for Genomic Medicine	74,236,262
	Luo	Yuling	Alamar Biosciences, Inc.	243
	Lyssand	John	NanoString® Technologies	102

	Last Name	First Name	Organization / Company	Page(s)
M	Maier	Keith	EpiCypher Inc.	152
	Mansfield	James	Visiopharm	292
	Marosy	Beth	Johns Hopkins University	25,148
	Marquis	Julien	University of Lausanne	219
	Martinez	Delsy	University of California San Francisco	226
	Masaeli	Maddison	Deepcell	182,246,271
	Masilionis	Ignas	Memorial Sloan Kettering Cancer Center	44,216 98
	Mehta	Monika		
	Miller	Katherine	University of Oxford	126,179,203,236
	Miller	Anthony	Institute for Genomic Medicine	74,126,179 203,236,262
	Minich	Jeremiah	Salk Institute for Biological Studies	110
	Mitchell	Jason	Delfi Diagnostics	33
	Moffat	Kelly	CosmosID	121
	Mohammed	Abdul	Volta Labs, Inc.	186
	Mohr	David	Johns Hopkins University,	25,148
	Monahan	Jack	biomodal Ltd	177
	Montano	Carolina	The Johns Hopkins University	25,104,148
	Moore	Richard	Michael Smith Genome Sciences Centre	255
	Morina	Luke	Johns Hopkins University	25,68,104,148 247,279
	Morrison	Carolyn	10x Genomics	269
	Mungall	Andre	Canada's Michael Smith Genome-Sciences Centre	52,255
N	Nagendran	Monica	10X Genomics	300
	Naishadham	Gautam	New England Biolabs	54,71,101
	Nakamoto	Maggie	Scale Biosciences	196,270
	Nakashian	Gregory	Broad Institute of MIT and Harvard	151
	Navarro	Jason	Abigail Wexner Research Institute at Nationwide Children's Hospital	179,236
	Nelen	Marcel	Pharmacy and Biomedical Genetics, UMC	97,218
	Newman	Martin	BioSkryb Genomics	140
	Ng	Kristie	Ontario Institute for Cancer Research	50
	Nicod	Jerome	The Francis Crick Institute	294

	Last Name	First Name	Organization / Company	Page(s)
O	Oguibe	God'salvation	University of Texas at San Antonio	165
	Oikonomopoulos	Spyros	McGill University	154
	Okada	Hiroyuki	The University of Tokyo	223
	Olson	Nathan	National Institute of Standards and Technology	77,86
	Oza	Andrea	The Broad Institute of MIT and Harvard	95,171
P	Park	Jiwoon	Weill Cornell Medicine	84
	Parnaby	Gavin	Illumina Inc	72
	Patrick	Michael	NanoString® Technologies	108
	Patterson	Taylor	Integrated DNA Technologies, Inc.	35
	Patterson	Ben	Vizgen	61
	Pel	Joel	Natera, Inc.	285
	Penkler	Jo-Anne	Roche Diagnostics Solutions	238
	Pinet	Kaylinnette	New England Biolabs	54,253
	Ponnaluri	Chaithanya	New England Biolabs	71,113,139,189 249,251,253
	Prasad	Nripesh	Discovery Life Sciences	174
	Pratto	Florencia	Genetics and Biochemistry Branch	118
Q	Qian	Xiaoyan	10x Genomics	170
	Quail	Michael	Wellcome Sanger Institute	144,214
R	Raeisi Dehkordi	Siavash	University of California San Diego	80
	Ragoussis	Jiannis	McGill University Genome Centre	154,220
	Raimer Young	Heather	New England Biolabs	54,71,249
	Rajan	Diana	Wellcome Sanger Institute	291
	Redshaw	Nicholas	Wellcome Sanger Institute	144
	Rhodes	Michael	NanoString® Technologies	137,282
	Riva	Alberto	University of Florida	64
	Rock	Rachel	Discovery Life Sciences	174
	Rodriguez	Deyra N.	New England Biolabs	54
	Rogert	Cande	Illumina Inc.	178,221
	Rojas	Alison	BioSkryb Genomics	198
	Rosenbloom	Alyssa	Nanostring Technologies, Inc.	102,129,137,297
	Rosenfeld	Jeffrey	Tonix Pharmaceuticals	117
	Roy	Sharmili	Stanford University	191

	Last Name	First Name	Organization / Company	Page(s)
S	Sakofsky	Cynthia	BD Biosciences	141
	Salek	Mahyar	Deepcell, Inc.	182,271,272
	Salman	Areesha	The Broad Institute of MIT and Harvard	171
	Sanders	Travis	Watchmaker Genomics	134,287
	Saunders	Lauren	Ceres Nanosciences	133
	Schmitz	Daniel	Uppsala University	65
	Scoones	Anita	Earlham Institute	172
	Sebra	Robert	Icahn School of Medicine at Mount Sinai	233
S	Sen	Moen	BD Biosciences	227
	Sepulveda	Hugo	La Jolla Institute for Immunology	58
	Sexton	Brittany	New England Biolabs Inc.,	43,49,54,252
	Shah	Shreyas	Revvity	164
	Shibata	Tatsuhiro	The University of Tokyo	46
	Shimazawa	Kei	The University of Tokyo.	83
	Short	Kirsty	University of Queensland	122
	Smith	Joshua	University of Washington	100
	Smith	Timothy	U.S. Meat Animal Research Center	119
	Smith	Owen	Twist Bioscience	185
	Soumillon	Magali	Flexomics	132
	Star	Ekaterina	Integrated DNA Technologies, Inc.	168,215
	Steemers	Frank	Scale Biosciences	270
	Stewart	Caitlin	Gencove	173
	Streck	Chris	n6 Tec, Inc.	226,264
	Sun	Zhiyi	New England Biolabs Inc.	201
	Sun	Yijia	Vizgen Inc.	281
	Suzuki	Ayako	The University of Tokyo	57
T	Teng	Dayu	UC San Diego	280
	Theall	Sandra	Roche	288
	Timp	Winston	Johns Hopkins University	68
	Tran	Miles	Brigham and Women's Hospital	99
	Trefry	Peter	Broad Clinical Labs	184

	Last Name	First Name	Organization / Company	Page(s)
U	Umapathi	Udayan	Volta Labs, Inc.	186
V	van der Zwaag	Albertus	UMC Utrecht	218
	Van Eijsden	Rudy	NanoString® Technologies	209
	Vats	Pankaj	Nvidia	67
	Venters	Bryan	EpiCypher	152,189,248
	Vieceli	John	Quantum-Si	163
	Villeneuve	Daniel	United States Environmental Protection Agency	109,115
	Vollbrecht	Thomas	Deepcell	246,271,272
W	Walker	Julie	Watchmaker Genomics	145
	Walsh	Cole	Broad Institute	217,286
	Walter	Dagmar	10X Genomics	92
	Wang	Kaile	UT MD Anderson Cancer Center	207
	Wang	Alexandra	Integrated DNA Technologies	250
	Wang	Mark	Illumina, Inc.	301
	Watson	Andrew	New York Genome Center	162
	Webber	James	Broad Institute	199
	Wicks	Andrew	New York Genome Center	267
	Williams	Louise	New England Biolabs, Inc.	54,71,113,139,249 251,252
Y	Ye	Rui	UT MD Anderson Cancer Center	38
Z	Zawistowski	Jon	BioSkryb Genomics	40
	Zhao	Junhua	Element Biosciences	254,302
	Zilionis	Rapolas	Droplet Genomics	162,213
	Zirkin	Shahar	JaxBio Technologies LTD.	263
	Zolj	Sanda	New England Biolabs Inc.	71
	Zong	Nan	ETH Zurich	125
	Zook	Justin	National Institute of Standards and Technology	41,77,86
	Zwirko	Zac	seqWell Inc.	131

WATCHMAKER GENOMICS

JOIN US | WEDNESDAY, 4:05 PM | PALMS BALLROOM II & III

FLIPPING THE SCRIPT ON DNA METHYLATION SEQUENCING

A positive readout to enable multimodal analyses

Discover Watchmaker's next focus area in our quest to maximize data recovery and integrity from challenging samples in sensitive applications. Can't wait? Stop by Bonaire 3 to learn more about our:

DNA Library Prep Solutions that address all sample types while yielding high conversion rates, reduced bias and artifacts, and reproducible results

RNA Library Prep Solutions that deliver superior sensitivity with FFPE and low-input samples and enable library prep in < 5 hours, including mRNA enrichment and rRNA depletion

Equinox Library Amplification Kits that provide excellent fidelity, uniform sequence coverage, and high library complexity for sensitive applications

Stop by Bonaire 3 for some sweet swag!

Don't miss Watchmaker's B*tchin' 80s Party

Be a boss and get your glow on!

Bonaire 3 | Tuesday 9:30 pm - 2 am

watchmakergenomics.com/AGBT24

PacBio

Welcome to tomorrow

PACB.COM/AGBT24

n6's iconPCR for NGS Library Prep
The Worlds First qPCR System with 96 Independently Controlled wells

Visit our Suite: Curacao 6!

Auto-normalization of NGS libraries
- Amplify each sample to the precise level - uniform yield from varying input
- Optimal for low input and challenging samples - cfDNA, single-cell, FFPE, ancient DNA
- Pre-pool samples for a single-tube bead cleanup-up to 90% cost reduction

Improved sequencing quality
- Prevent PCR artifacts - full length 165 and isomers with fewer chimeras
- Prevent over and under-amplification - drastic reduction in adapter dimers
- Intra-run quality control assessment using n6's novel QCurve settings

Allows unprecedented flexibility for assay optimization in a SINGLE RUN
- Create individual zones and truly precise temperature gradients
- Well-specific temperature differences for individual steps
- Variable PCR cycles for single shot DOE experimental design

Streamlined integration in any lab setting - no additional instrumentation
- Uses standard consumables, kits and assays- no specialty reagents required
- Automation compatible with standard liquid handling robots

SCAN ME

Learn more:
www.n6tec.com/AGBT2024
Poster 623, Wednesday, February 7,
4:45 P.M. - 6:10 P.M.