
AGBT2023™
February 6-9, 2023 • Hollywood, FL


**NanoString helps you see more biology than ever before.**

With the most comprehensive, fully integrated spatial solution set for collaboration, scalability, and flexibility, you can make your impact in spatial biology now.

**See it all on our Wall of Wow in Hospitality Suite 220**

AGBT2023™

GENERAL MEETING

Diplomat Beach Hotel • February 6-9, 2023

We welcome you to the 2023 **Advances in Genome Biology and Technology (AGBT) meeting.**

Now in its 23rd 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 (March 27-29, San Antonio, Texas) and AGBT Precision Health (September 7-9, San Diego, California)***

# 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 AGBT23 a content-rich experience.

**Penelope Bonnen**
*Baylor College of Medicine*

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

**Kim Doheny**
*Johns Hopkins University School of Medicine*

**Stacey Gabriel**
*Broad Institute of MIT and Harvard*

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

**Martin Hirst**
*University of British Columbia*

**Christopher Mason**
*Weill Cornell Medicine*

**John McPherson**
*University of California, Davis*

**Stephen Montgomery**
*Stanford University School of Medicine*

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

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

**Michael Zody**
*New York Genome Center*

# AGBT2023™

*We gratefully acknowledge the generous support provided by the following companies*

## GOLD

nanoString®

## SILVER

illumina®

10x GENOMICS

PacBio

## BRONZE

NEW ENGLAND BioLabs Inc.

Element Biosciences

BD

resolve biosciences

Parse BIOSCIENCES

BECKMAN COULTER Life Sciences

Roche

WATCHMAKER GENOMICS

SINGULAR GENOMICS

## CONTRIBUTING SPONSORS

BioSkryb GENOMICS

IDT INTEGRATED DNA TECHNOLOGIES

Complete GENOMICS Powered By MGI

AKOYA BIOSCIENCES

TWIST BIOSCIENCE

BRUKER

CODEXIS®

molecularloop

ULTIMA GENOMICS

BIO-RAD

TECAN.

ACD a biotechne brand

PARAGON GENOMICS

vizgen

Agilent Trusted Answers

bionano GENOMICS

PIXELGEN TECHNOLOGIES

ECLIPSEBIO THE RNA GENOMICS COMPANY

deepcell

PerkinElmer For the Better

QIAGEN OEM by QIAGEN

CHRYSALIS BIOMEDICAL ADVISORS

seqWell

DNASCRIPT

SCALE biosciences CELL TO INSIGHT

ROSALIND

# General Information

## Registration / Hospitality Desk

| | |
|---|---|
| Sunday, February 5th | 12:00 PM – 7:00 PM |
| Monday, February 6th | 8:00 AM – 7:00 PM |
| Tuesday, February 7th | 8:00 AM – 7:00 PM |
| Wednesday, February 8th | 8:00 AM – 7:00 PM |
| Thursday, February 9th | 8:00 AM – 6:00 PM |

### Name Badges

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

### Transportation

The Diplomat Beach Resort in Hollywood, Florida, is located 14 miles from Fort Lauderdale International Airport (FLL). The transfer from FLL is approx. 16 to 20 minutes (pending traffic). AGBT will offer complimentary shuttle service for arrivals and departures. Arrival Shuttles: Sunday, February 5th (1:00 PM – 8:00 PM) & Monday, February 6th (7:00 AM – 4:00 PM). Departure Shuttles: Friday, February 10th (4:00 AM – 2:00 PM).
Note: If you are arriving or departing on an alternate date/time, you will be responsible for securing your own transportation to the airport.

### Social Events

AGBT Welcome Reception – **Monday, February 6th 7:15 PM – 10:00 PM** at South Lagoon Pool, Playa, Infinity Pool, South Palm.

AGBT and Sponsor Dine Around - **Tuesday, February 7th 5:15 PM – 7:15 PM** South Palm Court & South Lagoon Pool. Dine around is a fun way to meet, mingle, and enjoy different light bites at multiple locations.

Women's Networking Event – **Tuesday, February 7th 5:30 PM – 7:15 PM** on 33rd Floor Lounge.

Click2Win – **Tuesday, February 7th 9:30 PM** in the Sponsor Promenade. Grand Prize will be awarded on Thursday's Plenary Session. (This prize is non-transferable.)

AGBT Dinner – **Wednesday, February 8th 6:15 PM – 7:25 PM** at South Lagoon Pool, South Palm Court & Monkitail Terrace

AGBT Farewell Dinner – **Thursday, February 9th 7:00 PM – 11:59 PM** at Portico at Diplomat Landing.

### Conference Technology

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

**#AGBT23**

# AGBT2023™
## AGRICULTURE

**MARCH 27-29, 2023**

SAN ANTONIO, TEXAS


REGISTER NOW FOR AGBT AG
TAKING PLACE NEXT MONTH.

REGISTER HERE

# Decades of innovation fueling discoveries

## NovaSeq™ X

See it for yourself in the Illumina lounge

illumina® | This is the Genome Era™

# Agenda

| | |
|---|---|
| **Sunday, February 5** | Transfers 1 p.m. – 8 p.m. Fort Lauderdale - Hollywood International Airport (FLL) |
| 12:00 p.m. – 7:00 p.m. | **Meeting Registration / Hospitality Desk**<br>Grand Foyer |
| **Monday, February 6** | Transfers 7 a.m. – 4 p.m. Fort Lauderdale - Hollywood International Airport (FLL) |
| 8:00 a.m. – 7:00 p.m. | **Meeting Registration / Hospitality Desk**<br>Grand Foyer |
| 10:00 a.m. – 4:00 p.m. | **Pre-Conference**<br>Grand Ballroom |
| 10:00 a.m. – 11:00 a.m. | **McDonnell Genome Institute, Washington University Ting Wang, Benedict Paten, Karen Miga**<br>Human Pangenome Reference Consortium (HPRC) Workshop |
| 1:00 p.m. – 4:00 p.m. | **Bionano Genomics**<br>**Adam Smith, Alexander Hoischen, Klint Rose, George Vacek, Erik Holmlin, Alka Chaubey, Alex Hastie, Mark Oldakowski, Marc Meyers, Damla Senol Cali, Sheila Purim**<br>"It's time we all move forward. Join the Bionano Pre-Conference Workshop to see how." |
| | **Scientific Program** |
| **Plenary Session:** | **Genomics I (Eric Green, National Human Genome Research Institute, Chair)**<br>Grand Ballroom |
| 5:00 p.m. – 5:10 p.m. | **Opening Remarks** |
| 5:10 p.m. – 5:40 p.m. | **Dame Sue Hill, Chief Scientific Officer, NHS England**<br>"Harnessing the potential of genomic medicine in the NHS in England" |

| | |
|---|---|
| 5:40 p.m. – 6:10 p.m. | **Pardis Sabeti, Broad Institute of MIT and Harvard**<br>"Sentinel: pandemic preemption in the genomic and information age." |
| 6:10 p.m. – 6:30 p.m. | **Christopher Mason, (Abstract Selected) Weill Cornell Medicine**<br>“The Spatial Atlas of Human Anatomy (SAHA) Project: A High content spatial map of the human body” |
| 6:30 p.m. – 6:50 p.m. | **Varsha Rao, (Abstract Selected) Claret Bioscience**<br>“Low-depth sequencing of cell-free DNA reveals disease-specific chromatin reorganization states” |
| 6:50 p.m. – 7:10 p.m. | **Hanlee Ji, (Abstract Selected) Stanford University**<br>"Single cell discovery of cancer point mutations and rearrangements with adaptive nanopore sequencing of transcripts" |
| 7:15 p.m. – 10:00 p.m. | **Welcome Reception – Havana Nights**<br>South Lagoon Pool, Playa, Infinity Pool, South Palm Court |
| **Tuesday, February 7** | **(Morning)** |
| 7:30 a.m. – 9:00 a.m. | **Breakfast**<br>South Palm Court |
| 8:00 a.m. – 7:00 p.m. | **Meeting Registration / Hospitality Desk**<br>Grand Foyer |
| **Plenary Session:** | **Evolutionary Genomics (Beth Shapiro, University of California, Santa Cruz, Chair, Chair)**<br>Grand Ballroom |
| 9:00 a.m. – 9:30 a.m. | **Love Dalen, Centre for Paleogenetics, Stockholm, University**<br>“A million-year genomic history of mammoth evolution” |
| 9:30 a.m. – 10:00 a.m. | **Carolyn Hogg, The University of Sydney**<br>“Conservation genomics: innovations, applications & where to next?” |

| | |
|---|---|
| 10:00 a.m. – 10:25 a.m. | **Elinor Karlsson, (Abstract Selected) UMass Chan Medical School**<br>"Investigating the genetics of complex diseases in thousands of dogs through community science" |
| 10:25 a.m. – 11:00 a.m. | **Coffee Break**<br>Sponsor Promenade |
| 11:05 a.m. – 11:25 a.m. | **Alexander Urban, (Abstract Selected) Stanford University**<br>"Automatic detection of complex structural variation across diverse human populations and their enrichment within genes of neurodevelopmental pathways" |
| 11:25 a.m. – 11:49 a.m. | **Poster Flash Talks** |
| **Tuesday, February 7** | **(Afternoon)** |
| 12:00 p.m. – 1:30 p.m. | **Gold Sponsor- NanoString Workshop & Lunch**<br><br>**Jennifer Hart, Director, Technical Sales – NanoString Technologies**<br>**Holger Heyn, PhD, Single Cell Genomics, Team Leader – Centre For Genomic Regulation, Spain**<br>**Vikram Devgan, PhD, Senior Director, Product Management – NanoString Technologies**<br>**Joseph Beechem, PhD, Chief Scientific Officer and Senior Vice President of Research and Development – NanoString Technologies**<br>**Miranda E. Orr, PhD, Assistant Professor – Wake Forest University School of Medicine.**<br><br>"The power of NanoString spatial biology is here"<br>Great Hall 4-6 |
| 12:00 p.m. – 1:30 p.m. | **AGBT Lunch**<br>South Palm Court |
| 1:30 p.m. – 3:30 p.m. | **Poster Session with Coffee & Dessert**<br>Regency Ballroom |

| | |
|---|---|
| **Plenary Session:** | **Genomics II (John McPherson, University of California, Davis, Chair)**<br>Grand Ballroom |
| 3:30 p.m. - 4:00 p.m. | **Benedict Paten, University of California, Santa Cruz**<br>"Building the human pangenome reference" |
| 4:00 p.m. – 4:30 p.m. | **Duncan MacCannell, Centers for Disease Control and Prevention (CDC), Office of Advanced Molecular Detection (OAMD)**<br>“Building sustainable pathogen genomic surveillance: Lessons from the COVID-19 Response” |
| 4:30 p.m. – 4:50 p.m. | **Billy Lau, (Abstract Selected) Stanford University School of Medicine**<br>“Nanopore sequencing of cell-free DNA for methylation-based breast cancer detection in a case-control cohort” |
| 4:50 p.m. – 5:10 p.m. | **Mitchell Vollger (Abstract Selected) University of Washington School of Medicine**<br>“Comprehensive diploid genetic and epigenetic profiles with single-molecule precision” |
| **Tuesday, February 7** | **(Evening)** |
| 5:15 p.m. - 7:15 p.m | **AGBT and Sponsor Dine Around**<br>South Palm Court, South Lagoon Pool, &<br>Sponsor Promonade<br><br>*Dine around is a fun way to meet, mingle, and enjoy different light bites at multiple locations.* |
| 5:30 p.m. – 7:15 p.m. | **Women’s Networking Event**<br>33rd Floor Lounge |

**Concurrent Session:**

| | |
|---|---|
| 7:30 p.m. – 9:30 p.m. | **Cancer (Stacey Gabriel, Broad Institute of MIT and Harvard, Chair)**<br>Grand Ballroom |
| 7:30 p.m. – 7:50 p.m. | **Edwin Cuppen, Hartwig Medical Foundation**<br>"Pan-cancer whole genome comparison of primary and metastatic solid tumors" |
| 7:50 p.m. – 8:10 p.m. | **John Hickey, Stanford University**<br>"Mechanisms of metaplastic progression to adeno-carcinoma revealed by high-speed multiomic spatial phenotyping of FFPE human samples" |
| 8:10 p.m. - 8:30 p.m. | **Amanda Janesick, 10x Genomics**<br>"High-resolution mapping of the breast cancer tumor microenvironment using integrated single cell, spatial and in situ analysis of FFPE tissue" |
| 8:30 p.m. – 8:50 p.m | **Julia Kennedy-Darling, Akoya Biosciences**<br>"High-speed multiomic spatial phenotyping of FFPE human cohorts to dissect mechanisms of environmental-conditioned metastasis" |
| 8:50 p.m. – 9:10 p.m. | **Rendong Yang, Northwestern University**<br>"Super-enhancer analysis identifies IGF1R-AS1 as a novel oncogenic LncRNA in advanced prostate cancer" |
| 9:10 p.m. – 9:30 p.m. | **Netanel Loyfer, The Hebrew University of Jerusalem**<br>"Deep whole-genome bisulfite sequencing of circulating cell-free DNA for monitoring patients with congestive heart failure" |

**Concurrent Session:**

| | |
|---|---|
| 7:30 p.m. – 9:30 p.m. | **Computational Biology (Mike Zody, New York Genome Center, Chair)**<br>Great Hall 4-6 |
| 7:30 p.m. – 7:50 p.m. | **Niall Lennon, Broad Institute of MIT and Harvard**<br>"Ancestry calibration of PRS population variance using AoU participant data" |

| | |
|---|---|
| 7:50 p.m. – 8:10 p.m. | **Victoria Popic, Broad Institute of MIT and Harvard**<br>"A deep learning framework for structural variant discovery and genotyping" |
| 8:10 p.m. – 8:30 p.m. | **Michael Schatz, Johns Hopkins University**<br>"Genomics at scale with the NHGRI AnVIL" |
| 8:30 p.m. – 8:50 p.m. | **Valerie Schneider, NIH/NLM/NCBI**<br>"The NIH Comparative Genomics Resource: A new ecosystem facilitating reliable comparative genomics" |
| 8:50 p.m. – 9:10 p.m. | **Yutaka Suzuki, The University of Tokyo**<br>"Intensive single cell analysis for elucidating immune cell diversity among healthy individuals" |
| 9:10 p.m. – 9:30 p.m. | **Jiannis Ragoussis, McGill University**<br>"Identification and characterization of SARS-Cov-2 defective interfering particles using direct RNA nanopore sequencing" |
| Starting at 9:30 p.m. | **Click2Win**<br>**Sponsor Promenade** |
| **Wednesday, February 8** | **(Morning)** |
| 7:30 a.m. – 9:00 a.m. | **Breakfast**<br>South Palm Court |
| 8:00 a.m. – 7:00 p.m. | **Hospitality Desk**<br>Grand Foyer |
| **Plenary Session:** | **Genetics (Kim Doheny, Johns Hopkins University School of Medicine, Chair)**<br>Grand Ballroom |
| 9:00 a.m. – 9:30 a.m. | **Alex Cagan, Wellcome Sanger Institute**<br>"The Impossibility of Whales: somatic evolution across the tree of life" |
| 9:30 a.m. – 10:00 a.m. | **Solenne Correard, The University of British Columbia**<br>"Building the path to equitable genomics care for Indigenous patients in Canada" |

| | |
|---|---|
| 10:00 a.m. – 10:30 a.m. | **Alicia Martin, Mass General Research Institute and Harvard Medical School**<br>"Genomics for the world: a comprehensive framework for genetic studies in diverse populations" |
| 10:30 a.m. – 11:10 a.m. | **Coffee Break**<br>Sponsor Promenade |
| 11:15 a.m. – 11:37 a.m. | **Poster Flash Talks** |
| **Wednesday, February 8** | **(Afternoon)** |
| 12:00 p.m. – 2:15 p.m. | **Silver Sponsor Workshops and Lunch**<br>Great Hall 4-6 |
| 12:00 p.m. – 1:00 p.m. | **Silver 1 Sponsor Workshop and Lunch – Illumina**<br>**Niall Lennon, PhD, Senior Director of Translational Genomics, Institute Scientist, The Broad Institute of MIT and Harvard**<br>**Alex Aravanis, Chief Technology Officer, Head of Research and Product Development**<br>**Fiona Nohilly, Sr Manager, Product Marketing, Illumina**<br>"NovaSeq X and innovations across the Illumina portfolio" |
| 1:10 p.m. – 2:10 p.m. | **Silver 2 Sponsor Workshop and Lunch —10X**<br>**Augusto Tentori, Malte Kühnemund, Sarah Taylor, and Jill Herschleb, 10x Genomics**<br>"Empowering impactful science with leading edge single cell, spatial, and in situ technologies" |
| 12:00 p.m. – 2:00 p.m. | **AGBT Lunch**<br>South Palm Court |
| 2:20 p.m. – 4:40 p.m. | **Bronze Sponsor Workshops**<br>Grand Ballroom<br>**(Stephen Montgomery, Stanford University School of Medicine, Chair)** |

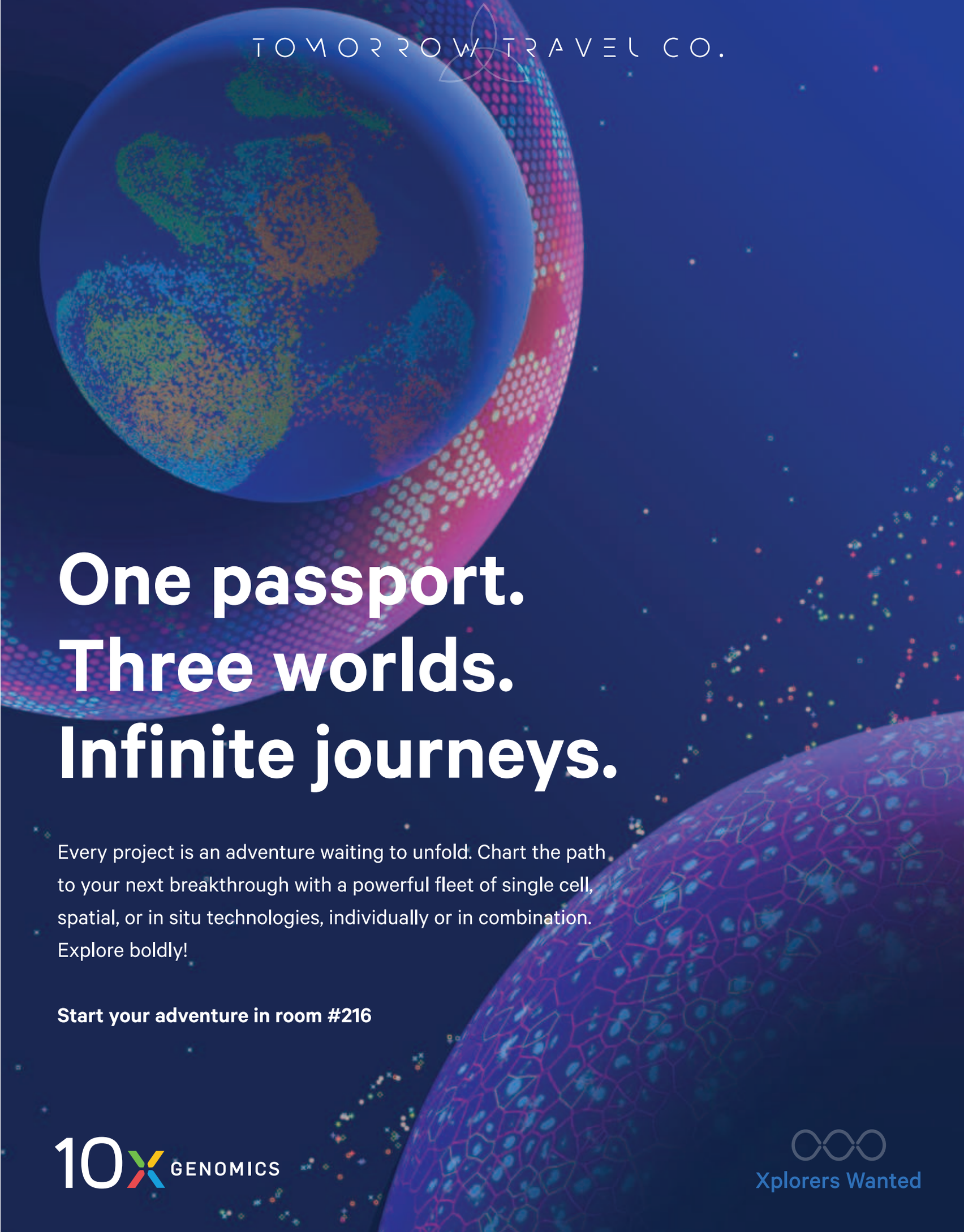
| | |
|---|---|
| 2:20 p.m. – 2:40 p.m. | **Bronze 1 Sponsor – New England Biolabs**<br>**Kit Krishnan, Ph.D., Development Group Leader, Next Generation Sequencing, New England Biolabs**<br>**Maggie Heider, New England Biolabs**<br>"Novel enzyme solutions enabling robust, high-performance library preparation from challenging to high-quality samples" |
| 2:40 p.m. – 3:00 p.m. | **Bronze 2 Sponsor – Element Biosciences**<br>**Matt Kellinger, PhD., Co-Founder & VP Biochemistry**<br>"Cloudbreak and Beyond: A roadmap for innovation" |
| 3:00 p.m. – 3:20 p.m. | **Bronze 3 Sponsor – BD Biosciences**<br>**Aruna Ayer, PhD, Director Single-Cell R&D, BD Biosciences**<br>"Single source for single-cell multiomics" |
| 3:20 p.m. – 3:35 p.m. | **Bronze 4 Sponsor – Resolve Biosciences, Inc**<br>**Jeroen Aerts, PhD**<br>"Spatial transcriptomics without compromising data quality" |
| 3:35 p.m. – 3:50 p.m. | **Bronze 5 Sponsor – Parse Biosciences**<br>**Charlie Roco, PhD, Cofounder, CTO**<br>"Broadening Access to Scalable Single Cell Sequencing" |
| 3:50 p.m. – 4:05 p.m. | **Bronze 6 Sponsor – Beckman Coulter Life Sciences**<br>**Ros Jackson, Ph.D., Director Scientific Research, Illumina, Calvin Cortes, MPH, MBA, Biomek NGeniuS Senior Product Manager, Beckman Coulter Life Sciences**<br>"Harnessing the power of automation: Illumina DNA Prep Kit + Biomek NGeniuS Next Generation Library Prep System" |
| 4:05 p.m. – 4:20 p.m. | **Bronze 7 Sponsor – Roche Sequencing and Life Science**<br>**Mariana Fitarelli Kiehl, PhD., Senior Applications Scientist, Roche Sequencing & Life Science**<br>"Single-tube total nucleic acid library preparation for FFPE: A combined method for DNA and RNA" |

| | |
|---|---|
| 4:20 p.m. – 4:32 p.m. | **Bronze 8 Sponsor – Watchmaker Genomics**<br>**Heather Maasdorp, Director of Global Support**<br>"Advancing NGS library preparation using cutting-edge enzyme engineering approaches" |
| 4:32 p.m. – 4:44 p.m | **Bronze 9 Sponsor – Singular Genomics**<br>**Eli Glezer, CSO, Founder**<br>"Rapid SBS on the G4 sequencer: expanded capabilities and new applications" |
| 4:45 p.m. – 6:10 p.m | **Poster Session and Wine & Cheese Reception**<br>Regency Ballroom |
| **Wednesday, February 8** | **(Evening)** |
| 6:15 p.m. – 7:25 p.m. | **AGBT Dinner**<br>South Lagoon Pool, South Palm Court & Monkitail Terrace |
| 7:30 p.m. – 9:30 p.m. | **Concurrent Session: Technology (Chris Mason, Weill Cornell Medicine, Chair)**<br>Grand Ballroom |
| 7:30 p.m. – 7:50 p.m. | **Andrew Adey, Oregon Health & Science University**<br>"High-throughput, robust single-cell DNA methylation profiling with sciMETv2" |
| 7:50 p.m. – 8:10 p.m. | **Joseph Beechem, NanoString**<br>"Path to the holy grail of spatial biology: spatial single cell whole transcriptomes using ultra-high plex spatial molecular imaging" |
| 8:10 p.m. – 8:30 p.m. | **Brian Hilbush, Veranome Biosystems/Applied Materials**<br>"A highly robust LISH-LnR workflow for subcellular in-situ single-cell spatial transcriptomic profiling of human liver tissue" |
| 8:30 p.m. – 8:50 p.m. | **Kunal Jindal, Washington University School of Medicine**<br>"Multiomic lineage tracing to dissect fate-specific gene regulatory programs" |

| | |
|---|---|
| 8:50 p.m. – 9:10 p.m. | **Heonseok Kim, Stanford University**<br>"Highly parallelized single cell modeling and phenotyping of cancer mutations with transcript informed CRISPR engineering" |
| 9:10 p.m. – 9:30 p.m. | **Kaile Wang, MD Anderson Cancer Center**<br>"Archival single cell sequencing reveals persistent subclones over years to decades of DCIS progression" |
| 7:30 p.m. – 9:30 p.m. | **Concurrent Session: Biology (Penelope Bonnen, Baylor College of Medicine, Chair)**<br>Great Hall 4-6 |
| 7:30 p.m. – 7:50 p.m. | **Iraad Bronner, Wellcome Sanger Institute**<br>"Combined Oxford Nanopore and PacBio sequencing of M. musculus for de-novo telomere-to-telomere genome assembly" |
| 7:50 p.m. – 8:10 p.m. | **JangKeun Kim, Weill Cornell Medicine**<br>"Inspiration4: Comprehensive and integrative single cell multi-omics analysis of the immune system of SpaceX Inspiration4 mission crews" |
| 8:10 p.m. – 8:30 p.m. | **Cindy Lawley, Olink**<br>"Empowering genomics with proteomics: Therapeutic target and biomarker discovery leveraging large population cohorts like the UK Biobank" |
| 8:30 p.m. – 8:50 p.m. | **Shawn Levy, Element Biosciences**<br>"Avidity-based sequencing improves sensitivity and detection in metagenomic analysis" |
| 8:50 p.m. – 9:10 p.m. | **Lindsay Rizzardi, HudsonAlpha Institute for Biotechnology**<br>"Identification of cis regulatory elements in DLPFC from Alzheimer's donors using single cell multiomics" |
| 9:10 p.m. – 9:30 p.m. | **Robert Sebra, Icahn School of Medicine at Mount Sinai**<br>"Isoform-resolved transcriptome reveals thousands of novel genes in human preimplantation embryos" |

| **Thursday, February 9** | **(Morning)** |
|---|---|
| 7:30 a.m. – 9:00 a.m. | **Breakfast**<br>South Palm Court |
| 8:00 a.m. – 6:00 p.m. | **Hospitality Desk**<br>Grand Foyer |
| **Plenary Session:** | **Biology (Federica Di Palma, Genome British Columbia, Chair)**<br>Grand Ballroom |
| 9:00 a.m. – 9:30 a.m. | **Freda Miller, The University of British Columbia**<br>"Using single cell approaches to understand how growth factors and stem cells build the brain" |
| 9:30 a.m. – 10:00 a.m. | **Judith Mank, The University of British Columbia**<br>“Y chromosomes matter” |
| 10:00 a.m. – 10:30 a.m. | **Beth Shapiro, University of California, Santa Cruz** |
| 10:30 a.m. – 11:10 a.m. | **Coffee Break**<br>Sponsor Promenade |
| 11:10 a.m. – 11:30 a.m. | **Peter Park, (Abstract Selected) Harvard Medical School**<br>“Single-cell genome sequencing of human neurons identifies somatic point mutation and indel enrichment in regulatory elements” |
| 11:30 a.m. – 11:50 a.m | **Somasekar Seshagiri (Abstract Selected) AntlerA Therapeutics**<br>“Next-gen snake anti-venom development through genome-guided single cell BCR-sequencing” |
| 12:00 p.m. – 1:00 p.m. | **Silver 3 Sponsor PacBio Workshop and Lunch**<br>Great Hall 4-6<br>Christian Henry, CEO, PacBio<br>“The power of high accuracy sequencing” |

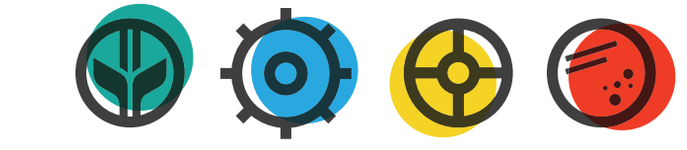
| | |
|---|---|
| 12:00 p.m. – 1:30 p.m. | **AGBT Lunch**<br>South Palm Court |
| **Thursday, February 9** | **(Afternoon)** |
| **Plenary Session:** | **Technology (Martin Hirst, University of British Columbia, Chair)**<br>Grand Ballroom |
| 1:30 p.m. – 2:00 p.m. | **Winston Timp, Johns Hopkins University**<br>"Beyond assembly: the increasing flexibility of single-molecule sequencing technology" |
| 2:00 p.m. – 2:30 p.m. | **Karen Miga, University of California, Santa Cruz**<br>"Expanding studies of global genomic diversity with complete, telomere-to-telomere (T2T) assemblies" |
| 2:30 p.m. – 3:00 p.m. | **Coffee Break**<br>Grand Foyer |
| 3:00 p.m. – 3:20 p.m. | **Andrew Russell, (Abstract Selected) Broad Institute of MIT and Harvard**<br>“Slide-tags: single-nucleus high-resolution multi-modal spatial genomics” |
| 3:20 p.m. – 3:40 p.m. | **Sanja Vickovic, (Abstract Selected) New York Genome Center**<br>“Spatial host-microbiome sequencing” |
| 3:40 p.m. – 4:00 p.m. | **Samantha Maragh (Abstract Selected), National Institute of Standards and Technology (NIST)**<br>"Outcomes of variant detection and quantitation from the first NIST Genome Editing Consortium Interlab Study" |
| 4:00 p.m. – 4:15 p.m. | **Click2Win Winner Announced**<br>**Closing Comments and Meeting Feedback** |
| 7:00 p.m. – Midnight | **Farewell Dinner – Glow Nights & After Glow**<br>Portico at Diplomat Landing |


TOMORROW TRAVEL CO.
One passport.
Three worlds.
Infinite journeys.
Every project is an adventure waiting to unfold. Chart the path to your next breakthrough with a powerful fleet of single cell, spatial, or in situ technologies, individually or in combination. Explore boldly!
Start your adventure in room #216
10X GENOMICS
Xplorers Wanted

# Abstract Table of Contents

page





AGBT2023™

PacBio


LONG READS **AND** SHORT READS

# Game, changed.

PACB.COM/AGBT23

# Cancer Omics

# A novel suite of enzyme mixes enable robust hybrid capture sequencing from low-quality FFPE DNA



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

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

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

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

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

# A pathway-centric approach to characterizing tumor heterogeneity and cell diversity across multiple cancer types



**Leiam Colbert[1], Ben Patterson[1], Cheng Yi Chen[1], Nicolas Fernandez[1], Jiang He[1]**

1 Vizgen, 61 Moulton St, Cambridge, Massachusetts, United States

Biological tissues are composed of heterogeneous populations of cells intricately organized into 3D architectures. Single-cell RNA sequencing provides a systematic and quantitative approach to characterizing cell type and composition within tissue. However, the spatial organization of cells and cell-cell interactions, which are critical for understanding tissue function, are lost when cells are dissociated from tissue for single-cell RNA sequencing. For oncologists, the spatial information embedded in the heterogeneous tumour microenvironment plays a critical role in assessing patient prognosis. Recently, spatially resolved, single-cell imaging platforms have provided a necessary solution to bridge the spatial information gap evident in previous transcriptomic methodologies. Vizgen's MERSCOPETM platform, built on Multiplexed Error-Robust Fluorescence in situ Hybridization (MERFISH) technology, enables the direct profiling of the spatial organization of patient tumours with subcellular resolution. Here, we present a Pan-Cancer approach to characterize various cancers in human clinical samples using the MERSCOPETM Platform. Specifically, using a 500-gene panel that includes the canonical signalling pathways of cancer, cancer type-specific genes, key immune genes, proto-oncogenes, and tumour-suppressor genes, we demonstrate MERSCOPE's ability to spatially profile gene expression across multiple tumour types, including breast, colon, prostate, ovarian, lung, and skin cancer. To confirm the specificity and sensitivity of our pathway-focused gene panel, we assessed how the tumour cell clusters in each different dataset expressed different cancer type-specific markers. As expected, we found specific cancer types were enriched for their corresponding marker genes. To deeply assess how individual cell types in each tumour are dysregulated by cancer, we compared non-tumour cell clusters to similarly annotated clusters derived from single-cell RNA-sequencing data of corresponding normal tissues. Notably, the non-tumour cells in these datasets exhibit higher expression of stress and inflammation pathways, while immune cell clusters demonstrate higher immune activity and higher immune exhaustion compared to those in normal tissue. These results demonstrate the power of the MERSCOPE platform to generate individualized, accurate cell atlases from patient-derived tumours and to enable further insights into the relationship between genomic profiles, dysregulated pathways, and disease phenotype. Such network-centric approaches are critical for identifying genotypic causes of diseases, classifying disease subtypes, and identifying drug targets.

# Accurate detection of epidermal clonal mutations for nonmelanoma skin cancer prevention



**Lei Wei**[1]

1 Department of Biostatistics and Bioinformatics, Roswell Park Comprehensive Cancer Center, Buffalo, NY

The most common human malignancy, non-melanoma skin cancer (NMSC), is predominantly caused by exposure to ultraviolet (UV) rays.

UV is known to introduce C>T transition or CC>TT dinucleotide mutations at dipyridine sites, referred to as UV-signature mutations (USMs). The accumulation of USMs and other secondary mutations fuels the epidermal clonal expansion in normal skin, leading to the progression to actinic keratosis (AKs) or even NMSCs. Characterizing these clonal mutations (CMs) in normal-looking skin would provide invaluable information about the skin's early photo-carcinogenesis and help develop better prevention strategies. Recent studies reported high CM burdens in the sun-exposed human epidermis but lacked the differentiation between UV and other factors such as aging or environmental factors. To elucidate the exact mutational effects of UV exposure on the skin, we performed targeted ultra-deep sequencing (UTS) of 50,000X in 450 individual-matched sun-exposed (SE) and non-sun-exposed (NE) epidermal biopsies obtained from clinically normal skin from 13 donors. A total of 638 CMs were identified, including 298 USMs. The USMs per sample were three times higher in the SE samples and were associated with significantly higher variant allele frequencies (VAFs), compared with the NE samples. Several genomic regions in TP53, NOTCH1 and GRM3 were specifically mutated in SE skin. We defined Cumulative Relative Clonal Area (CRCA), a single metric of UV damage calculated by the overall relative percentage of the sampled skin area affected by CMs. The CRCA was dramatically elevated by a median of 11.2 fold in SE compared to NE samples. In an extended cohort of SE normal epidermal samples from patients with a high- or low- burden of cutaneous squamous cell carcinoma (cSCC), the SE samples in high-cSCC patients contained significantly more USMs than SE samples in low-cSCC patients, with the difference mostly conferred by mutations from low-frequency clones (defined by VAF⩽1%) but not expanded clones (VAF>1%). Our two studies of mutational features in normal skin between paired SE/NE sites and high/low-cSCC risk patients revealed novel insights into the tumorigenesis progression after chronic UV exposure, and provided new opportunities for developing CM-based biomarkers to assess NMSC risk and optimize intervention strategies.

# Application of spatially resolved transcriptomics to screen multiple tumor biospecimens using tissue microarrays



**Cedric R. Uytingco[1], Syrus Mohabbat1, Hardeep Singh[1], Stephen R. Williams[1], Lauren M Gutgesell[1], David J Sukovich[1], Govinda M Kamath[1], Hanyoup Kim[1], Augusto M Tentori[1]**

1 10x Genomics, Pleasanton, CA

The tumor microenvironment is composed of highly heterogeneous structures and cell types that dynamically influence and communicate with each other. Although examination of singular biospecimens is sufficient for diagnostic purposes, it is inadequate and cost prohibitive when scaling for complex and overarching studies. Thus, high density multi-tumor tissue microarrays (TMAs) have been a practical and effective solution for high-throughput molecular analysis of tissues. Introduced more than a decade ago, TMAs have been instrumental in the recent study of tumor biology, the development of diagnostic tests, the establishment of quality control, and the investigation and identification of oncological biomarkers. Here, we demonstrate the application of the 10x Genomics Visium Cytassist Spatial Gene Expression Solution on archival multi-tumor TMAs to screen for common cancer genes among a cohort of breast cancer and multi-tumor samples. Spatial transcriptomics technology has proven valuable in complementing pathological examination by combining histological assessment with the throughput and deep biological insight of RNA-seq. Visium CytAssist expands the accessibility of the pre-existing standard Visium solution by facilitating the retrieval of RNA transcriptomic information from tissues placed on standard or archival slides. Supporting capture areas of 6.5 x 6.5 mm and 11 x 11 mm, RTL-based probes with spatial barcodes generate a high quality and sensitive mapping of the transcriptome. Through the Visium CytAssist workflow, we showcase the ability to spatially and comprehensively resolve individual oncogene and tumor suppressor genes associated with multiple tumors from a cohort of breast cancer patients and from multiple different tumor samples. In addition, these markers are mapped back to distinct morphological features within each tissue core, and use differential gene expression data to identify distinct cell types throughout the different patient tissues. Overall, these data highlight that Visium CytAssist can retrieve transcriptome information from archived FFPE multi-tumor TMA sections while preserving its spatial context. By combining the throughput of TMA samples and depth of the Visium platform, the strategy enables greater insights into cell-type specific while also expanding the spectrum of biospecimen types that can be analyzed.

# Assessment of the utility of a novel massively parallel sequencing platform for multiplexed somatic cancer assays

**Sophie Low[1], Michelle Cipicchio[1], Nicole Francis[1], Ning Li[1], Mark Fleharty[1], Micah Rickles-Young[1], Carrie Cibulskis[1], Stacey Gabriel[1], Niall Lennon[1]**

1 Genomics Platform, Broad Institute of MIT and Harvard, Cambridge, MA 02141

Tumor profiling using next generation sequencing can provide powerful information to guide patient diagnosis and care, however enabling flexibility in scale can be a challenge for core labs. Cancer specific panels are an important tool for both research and clinical testing. Projects vary significantly in scope, with variable sample numbers, target territory and sequencing depth required. Expanding into high sensitivity platforms using unique molecular index error correction requires low sequencing cost per base due to the very deep coverage required. Cancer research requires flexible scale, with throughput that may be too low to leverage the most cost effective sequencing options available. Despite substantial multiplexing capabilities, high-throughput sequencers do not often fit well with smaller scale clinical testing platforms that require fast turn-around times and the ability to process a few samples at a time. We are exploring a novel sequencing platform that can provide a solution to enable flexibility in scale while matching the most advantageous cost. With high quality, low error NGS reads of up to 150 bp, the Element Aviti benchtop sequencer provides a solution with flexible scale that would enable small batch sizes.

We present an analysis of a Pan Cancer gene profiling panel using deep sequencing on samples with somatic variants at known allele fractions. Additionally we assess the performance of ultra low coverage whole genome sequencing for tumor fraction estimation from liquid biopsy. We will evaluate this platform for workflow, cost effectiveness, and variant calling sensitivity and specificity. The low error rate on this platform, combined with the attractive per base cost and flexible scale, represents a solution for the complex challenges limiting the use of NGS in cancer today.

# Benefits of applying molecular barcoding systems are not uniform across different NGS-based applications



**Slawomir Kubik, Jonathan Bieler, Christian Pozzorini, Adrian Willig, Zhenyu Xu**

SOPHiA GENETICS, Data Science Department, Saint Sulpice, Switzerland

High-throughput sequencing of tumor DNA combined with target enrichment has the potential to facilitate personalized cancer therapy by accurately detecting mutations with increasingly lower allele frequencies. For example, circulating tumor DNA (ctDNA) and white blood cell-derived DNA can be analyzed to determine residual disease levels for solid tumors and lymphoid disorders, respectively. Detection of low-frequency, tumor-specific mutations is also relevant for formaldehyde-fixed paraffin-embedded (FFPE) specimens with low tumor content. Nevertheless, low quantities of ctDNA in the blood, the complexity and diversity of immunoglobulin and T cell receptor rearrangements in lymphocytes, and high levels of noise relative to the frequency of targeted mutations in FFPE samples all limit the analytical performance of rare variant detection.

Unique Molecular Identifiers (UMIs), constituting a molecular barcoding system, coupled with computational methods of noise suppression, are believed to drastically improve the reliability of genotyping. We demonstrate that UMI usage is not universally beneficial in every experimental setup. We dissect the major challenges related to low-frequency variant calling and highlight those that can be overcome by using molecular barcodes. We demonstrate, using data generated from various types and quantities of input material, that variant calling based on molecule counting via DNA fragment mapping positions, displays optimal performance in some scenarios. Exogenous barcodes provide significant improvement only when mapping position collisions occur frequently, for example in ctDNA, where nucleic acid fragmentation is non-random but rather driven by nucleosome positioning. We also present a novel UMI system, Spacer Length Adapters (SLAs), where template molecule identity is encoded by adapter size rather than its sequence. SLAs demonstrate comparable performance to existing UMI solutions, allow efficient suppression of noise, and facilitate rare variant detection.

Our findings provide guidance on whether the implementation of a UMI system for a given NGS application is advantageous.

# Cross-tissue comparison of protein-based spatial biology methods



**Huan Wang[1], Jia-Ren Lin[2,3], Sachi Krishna[1], Zoltan Maliga[2,3], Domenic Abbondanza[1], Yu-An Chen[2,3], Kathleen Pfaff[6], Anne E Carlisle[6], Peter K Sorger[2,3,4], Scott Rodig[5,6], Sandro Santagata[2,3,5], Samouil L Farhi[1]**

1 Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA.
2 Ludwig Center for Cancer Research at Harvard, Harvard Medical School, Boston, MA, USA.
3 Laboratory of Systems Pharmacology, Harvard Medical School, Boston, MA, USA.
4 Department of Systems Biology, Harvard Medical School, Boston, MA, USA.
5 Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA.
6 Center for Immuno-Oncology Tissue Biomarker Laboratory, Dana-Farber Cancer Institute, Boston, MA, USA.

Multiplexed antibody imaging enables investigations of spatial biology, including the structure of the tissue microenvironment, cell-cell interactions, and tumor-immune contacts. While several multiplexed imaging methods have been developed such as CO-detection by inDEXing
(CODEX), cyclic immunofluorescence (CyCIF), and highly multiplexed immunofluorescence (hmIF), no dataset exists in which these methods have been applied to the same set of samples. It thus remains unclear how cell type and cell state calling are influenced by the acquisition method.

To address this gap, we created a tissue microarray (TMA) consisting of 170 cores from seven cancer types: colorectal cancer (CRC), head and neck squamous cell carcinomas (HNSCC), melanoma, breast cancer (BrC), non-small cell lung cancer (NSCLC), and ovarian cancer (OV), and nine total tissue types, with tumor and normal pathologist annotation. Sequential sections were distributed to labs running highly multiplexed protein imaging platforms (PhenoCycler, CyCIF, and PhenoImager) to be profiled with validated antibody panels. We observed qualitatively similar staining patterns across techniques across different tissue types.

*(continue to page 32)*

*(continued from page 31)*

Data was then exchanged between the labs such that each data type was processed by each lab's analysis platform for stitching, registration, segmentation, and cell type calling. While the overall data processing steps were similar, each was accomplished by different algorithms and implementations. We then quantified signal-to-noise ratios for individual markers, agreement of automated cell type assignments with manual calling, and cell counts and composition by tissue type. With this data, we are able to evaluate how dependent results from such studies are on the techniques used to profile the tissue and the platforms used to analyze the data; and whether questions about particular tissues or cell type populations are better answered by different techniques.

# Driver panorama of 1,335 multi-ancestry gastric cancer genomes



**Tatsuhiro Shibata**

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

Gastric cancer is one of the most common malignancies worldwide, with geographic, epidemiological, and histological diversities. We report an extensive (1,335 cases), multi-ancestry landscape of driver events in gastric cancer. Seventy-seven significantly mutated genes were identified in total. We also highlight subtype-specific drivers including PIGR and SOX9, which were both enriched in the diffuse subtype of the disease. Significantly mutated genes also varied according to Epstein-Barr virus status and ancestry. Non-protein truncating CDH1 mutations characterized by in-frame splicing alterations targeted localized extracellular domains and uniquely occurred in sporadic diffuse-type cases. In cases of East Asian ancestry, our data suggest a link between alcohol consumption or metabolism with the characteristic diver mutations. Moreover, we identify mutations with potential roles in immune evasion. Overall, these data provide a comprehensive insight into the molecular landscape of gastric cancer across subtypes and ancestries.

# Genome in a Bottle mosaic benchmark for HG002: Initiating MDIC's somatic reference sample program



**Adetola Abdulkadir[1], Camille Daniels[2], Nathan D. Olson[3], Megan Cleveland[4], Maryellen de Mars[5], Justin Zook[6]**

1 Medical Device Innovation Consortium, Cancer Genomics Somatic Reference Samples Initiative, Arlington, VA, United States of America
2 Medical Device Innovation Consortium, Cancer Genomics Somatic Reference Samples Initiative, Arlington, VA, United States of America
3 National Institute of Standards and Technology, Material Measurement Laboratory, Gaithersburg, MD, United States of America
4 National Institute of Standards and Technology, Material Measurement Laboratory, Gaithersburg, MD, United States of America
5 Medical Device Innovation Consortium, Cancer Genomics Somatic Reference Samples Initiative, Arlington, VA, United States of America
6 National Institute of Standards and Technology, Material Measurement Laboratory, Gaithersburg, MD, United States of America

Cancer-causing somatic mutations are of interest to research and diagnostic communities for use as cancer biomarkers. The identification and characterization of such biomarkers are helping to make precision medicine a reality by more effectively diagnosing and applying targeted cancer treatments. The Medical Device Innovation Consortium (MDIC) identified a lack of well characterized somatic variant reference samples and benchmarks for use in development of cancer biomarker diagnostics. The Somatic Reference Samples (SRS) Initiative was launched by MDIC to address this need by engineering important somatic variants into the extensively characterized Genome in a Bottle (GIAB) sample HG002. To establish a baseline for the engineered cell lines, we first are identifying low frequency mosaic variants in HG002. Mosaic variants in the GIAB HG002 reference material DNA were identified using a new trio-based approach with the somatic variant caller strelka2. To determine if the strelka2 variant caller is fit for purpose, we used in-silico mixtures of real data from HG002 and HG003 to determine the allele frequency limit of detection. Then, high coverage (300X) Illumina data from the Ashkenazi Jewish trio was used for low frequency variant detection in the son. Data from HG002 (son) was passed to strelka2 as the tumor data and the combined parents' data (HG003-father and HG004-mother) was passed as normal. These candidate mosaic variants identified with strelka2 were subsequently evaluated using high-coverage PacBio HiFi data as well as TwinStrand duplex sequencing. A custom targeted duplex sequencing panel synthesized by Twist Biosciences included ~1200 potential mosaic variants and ~300 control variants that were known to be heterozygous in HG003 and homozygous reference for HG002.

*(continue to page 35)*

*(continued from page 34)*

The control variants are positive controls for HG003 samples and negative controls for HG002. Additionally, DNA mixtures of HG002 and HG003 were used to evaluate the quantitative accuracy of the duplex sequencing allele fraction measurements. The characterization of mosaic variants in HG002 samples inform what sequencing and informatic methods we will use to characterize the forthcoming somatic reference samples from the SRS Initiative.

# High-plex spatial multiomics of pediatric and adult diffuse midline glioma, H3 K27-altered



**Gary Geiss, Shawn Damodharan DO[1] and Mahua Dey MD[2]**

1 Department of Pediatrics, Division of Pediatric Hematology, Oncology and Bone Marrow Transplant, University of Wisconsin School of Medicine & Public Health, Madison, WI, USA
2 Department of Neurosurgery, University of Wisconsin School of Medicine & Public Health, UW Carbone Cancer Center, Madison, WI, USA

DMG, H3 K27-altered are aggressive malignancies of the brain that affect both pediatric and adult populations. The NanoString GeoMxTM Profiler and Illumina sequencing platforms was used on a cohort of both pediatric and adult samples. The analysis showed differences which may contribute to the clinical differences between both populations.

# In-situ estimation of T-cell receptor diversity and transcriptomics with NanoString Digital Spatial Profiling in FFPE tumor biopsies



**Li Fan[1], Mahesh Bachu[1], Manling Ma-Edmonds[1], Angelo Harris[2], Anantharaman Muthuswamy[2], Suzana Couto[2], Mark Fereshteh[1], Omar Jabado[1]**

1 Translational Sciences, Genmab, Princeton, NJ, USA
2 Pathology, Genmab, Princeton, NJ, USA

Building on the success of PD1 checkpoint inhibition in oncology, many T-cell modulation strategies are being tested in clinical trials. The core of these strategies is to drive a tumor-specific antigen response, enabling a robust and systemic tumor elimination by cytotoxic T cells. Evidence of such a response in tumor is difficult to obtain by T-cell receptor (TCR) sequencing with formalin-fixed paraffin-embedded (FFPE) tissue due to RNA degradation. Spatial biology platforms, such as the NanoString GeoMx Digital Spatial Profiler (DSP), have enabled high resolution transcriptomic studies of the tumor microenvironment. We used a new ~200 gene TCR spike-in panel combined the current whole transcriptome DSP panel to assess T-cell diversity by V/J gene usage in tumor microarrays (TMAs).

A total of 163 patients were assessed in the following cancer types: colorectal, ovarian, triple-negative breast, and head and neck. We used the exquisite targeting features of DSP to profile CD3+ cells and compared them to adjacent tumor and CD3- stroma with the goal of identifying which V/J probes were specific to T-cells. Probes for αβ chain genes showed high abundance and specificity (J 99% and V 89% detectable). Although γδ probes showed lower abundance, detection rate was still high (J 100% and V 71%). The number of T-cells in the tissue region profiled was proportional to diversity; genes highly corelated to diversity were associated with T-cell proliferation (CD52, ANP32B, CIITA). We applied a Jaccard similarity measure to compare gene usage across the cohort and found intra-patient samples were more similar in V/J usage than inter-patient, as expected. Further exploration of probe sensitivity, appropriate diversity statistics and comparison to gold standard TCR sequencing methods are underway.

This novel platform enables monitoring changes in T-cell diversity and transcriptome-wide profiling with spatial resolution in FFPE tumors. The resulting data may be useful in mechanism of action studies for immunotherapy development to quantify potentially antigen specific T-cells.

# Integrated short- and long-read single-cell sequencing shows differential IL32 isoform usage in tumor-infiltrating lymphocytes following immune checkpoint perturbation



**Susan Grimes**[1], **Anuja Sathe**[1], **Heonseok Kim**[1], **Hanlee Ji**[1]

1 Division of Oncology, Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, United States

Immunotherapy agents called immune checkpoint inhibitors block the binding of checkpoint proteins to reactivate a cell's immune response. However, the precise mechanism of action is not fully understood. We investigated the cellular and molecular responses to inhibition of the immune checkpoint receptor TIGIT, using ex vivo cultures from five surgical resections of colorectal and gastric cancer. These cultures maintain the patient's original tumor microenvironment (TME) in a near-native state. We used single cell RNA-seq with integrated short-reads and long-reads to define gene expression, cell type information and full transcript structure. This method revealed the complex cell interactions occurring in the multi-cellular TME.

From short-reads, significant up-regulation of the cytokine Interleukin 32 (IL32) was observed in both CD8 T-cells and CD4 TFh-like cells following treatment. IL32 has seven major alternatively-spliced isoforms that contribute to different functions including cell differentiation, apoptosis and pro or anti-inflammatory reactions. IL32 transcript isoforms were determined from long-read data, and deconvoluted to each single cell or cell type by matching short-read and long-read barcode sequences. The IL32β isoform was the most prevalently found in all samples, with moderate-low incidence of IL32θ, IL32γ, IL32α, IL32ζ and IL32δ in descending order of frequency, and very few cells with IL32ε. After treatment with TIGIT inhibitor, a significant shift to higher and more heterogeneous proportions of IL32β, IL32γ and/or IL32θ isoforms per cell was observed in CD8-T cells in the four samples that had a transcriptional response of increased cytotoxicity. A similar isoform pattern was observed in CD4-TFh cells in three of the four responding samples. There was no significant change in IL32 isoform usage in B cells, or in stromal cells for any sample.

The shift of transcripts between IL32β, IL32γ and IL32θ isoforms in immune cells has been previously reported and is posited to modulate inflammation, with IL32γ and IL32θ having the highest inflammatory activity. This effect could be important in mediating an immune activated TME following TIGIT inhibition. Overall, our approach allows the single cell investigation of cell type specific effects in patient-derived model systems in response to immunotherapy, potentially revealing novel genome biology..

# Integration of the ResolveOME multi-omics workflow with IDT xGen hybridization capture at single-cell resolution to obtain high quality exome data



**Jon S. Zawistowski[1], Durga M. Arvapalli[1], Isai Salas-González[1], Benjamin Warlick[1], Viren Amin[1], Victor J. Weigman[1], Gary L. Harton[1], Jay A.A. West[1].**

1 BioSkryb Genomics, Durham, NC.

Simultaneous examination of multiple-omics layers in single cells offers essential insights to decode the complexity of cancer and understand the pathophysiology of oncogenesis as well as tumor progression. ResolveOME TM amplification chemistry provides these insights through unified whole-genome and full-length transcriptome information from the same single cell. While whole genome sequencing (WGS) provides the complete complement of genomic elements for discovery, whole exome sequencing (WES) offers a cost-effective alternative while still providing the ability to ascertain the entirety of coding sequence.

Here we extend ResolveOME TM to a robust hybrid capture workflow to examine whole exomes. The core ResolveOME TM workflow is unification of template-switching single-cell RNAseq chemistry with ResolveDNA® whole genome amplification (WGA) technology, whereby the cDNA synthesized is separated from the genomic amplification fraction through affinity purification. The ResolveDNA® Library Preparation kit with enzymatic fragmentation was used to create libraries from the amplified genomic fraction. Prior to performing hybrid capture, low pass sequencing ensured high library complexity as assessed by the preseq algorithm. The libraries were pooled and enriched using the IDT xGen TM Exome Research Panel v2. The pool was 2X150 sequenced on an Illumina NextSeq 1000 instrument. The samples were downsampled to 40 million reads per library where analysis using BaseJumper TM software showed target base coverage of 98% at 1X and 96% at 10X. The single-cell SNV calling metrics showed excellent precision and sensitivity of 99% and 92% respectively with robust allelic balance across single cells and on target bases over 70%.

We demonstrated the successful coupling of the ResolveOME TM multi-omics workflow with IDT xGen TM Research Panel v2 hybridization capture of genomic libraries, providing the researcher a WES option in addition to WGS for maximizing cost and efficiency, with the key attributes of robust exomic metrics.

# Low-input multi-omics from laser microdissection

**Michael A. Quail,** **Scott Goodwin, Matt Mayho, Jyoti Nangalia and Inigo Martincorena**

Wellcome Sanger Institute, Hinxton, United Kingdom

Somatic mutations accumulate in all tissues as we age. They underpin cancer development and have been proposed to contribute to ageing and other diseases. Despite their importance, detecting somatic mutations in normal tissues remains challenging due to mutations being often present in microscopic clones of cells. Laser microdissection of a few hundred cells followed by low-input DNA sequencing has become a powerful method to study these mutant clones. At the Sanger Institute, we have established a high-throughput pipeline to perform high-quality genome or exome sequencing from a few hundred cells without whole-genome amplification prior to adapter ligation, which has been used on thousands of samples and multiple publications in the past 3 years. Here, we describe new protocols to carry out high-quality multiomic assays (DNA, RNA and DNA methylation) on microscopic areas of tissue.

# Mechanisms of therapeutic resistance at the tumor-stromal interface using spatial molecular imaging and genetically-engineered tumoroids



**William Hwang**

Massachusetts General Hospital

Pancreatic ductal adenocarcinoma (PDAC) is a highly lethal malignancy with a 5-year survival rate of 11%. Given its limited response to cytotoxic therapy, it is crucial to identify interactions at the tumor-stromal interface that mediate therapeutic resistance.

We performed single-cell spatial transcriptomics on 14 primary resected PDAC tumors (treatment-naive, n = 7; chemotherapy/radiotherapy, n = 7) with spatial molecular imaging (SMI; NanoString CosMx) using both standard (960 targets) and partially customized panels (990 targets including 30 custom spike-ins).

We profiled more than 700,000 high-quality cells, segmenting and annotating cells using a semi-supervised clustering algorithm leveraging previously generated single-nucleus RNA sequencing (snRNA-seq) expression profiles from 224,988 cells and 43 patient samples defining canonical cell type marker genes. By integrating the single-nucleus sequencing and the spatial molecular imaging data, we identified 19 distinct cell types, including epithelial, immune, endothelial, endocrine, and diverse stromal cell types and subtypes. We further dissected the malignant, fibroblast, and endothelial populations by applying snRNA-seq derived program signatures.

To explore the multicellular tumor architecture and cell-cell interactions, we investigated intra-glandular heterogeneity, multicellular neighborhood compositions, and treatment-enriched ligand-receptor interactions at single-cell spatial resolution. We computationally modeled a spatially-aware non-negative matrix factorization (sNMF) to extract distinct multicellular communities, and we developed a novel computational framework leveraging optimal transport to infer intercellular communication. We identified ligand-receptor interaction pairs that were associated with therapeutic resistance. We functionally validated a subset of these predicted interactions using a murine stromal tumoroid model derived by co-culturing transformed pancreatic ductal cells and cancer-associated fibroblasts (CAFs) in a solubilized basement membrane preparation (Matrigel), performing ex vivo chemotherapy treatment followed by RNA-seq.

*(continue to page 42)*

*(continued from page 41)*

We are now using CRISPR technology to modulate the expression of select ligand-receptor pairs in the malignant-CAF co-cultures and testing therapeutic resistance both ex vivo and in vivo with syngeneic orthotopic transplants.
By integrating snRNA-seq, high-resolution spatial transcriptomics, novel computational methods, genetically-tractable murine models, we have established a new paradigm to identify mechanisms of therapeutic resistance based on critical intercellular interactions in the tumor microenvironment of pancreatic cancer, which will be broadly applicable to other malignancies and diseases.

# Methylation controls to detect for methylation level quantification in the Twist Targeted Methylation Sequencing workflow



**Esteban Toro**, **Kristin Butcher, Twist Bioscience, Research & Development**

DNA methylation at CpG nucleotide sites in eukaryotes is a key epigenetic mark that helps control gene expression. Specific changes in CpG methylation occur in many human cancers, making them a promising biomarker for early cancer detection, especially in the context of liquid biopsy testing. However, assays for methylation detection are often only semi-quantitative for the methylated fraction, making it difficult to firmly establish the limit of detection for these assays and to define methylation levels in a given sequence.

To address these challenges, Twist Bioscience has designed and developed a CpG methylation specific control to help calibrate assays that determine site-specific methylation rates. By using our core DNA synthesis technology, we have constructed a pool of 48 unique contrived sequences that contain a total of 8 different levels of methylation at a CpG site, ranging from 100% to 0%. Each unique sequence also contains either a low, medium, or high amount of total CpG sites for direct comparisons of sequences with various CpG sites. These controls also mimic the DNA fragment lengths commonly found in cell-free DNA (cfDNA) and are therefore compatible with experimental workflows that target liquid biopsy applications. These pools can be taken through the Twist Targeted Methylation Sequencing workflow, using panels designed to target these controls as spike-in panels during hybrid capture.
This study demonstrates an improved method for calibrating detection assays across a range of methylation levels, containing variable amounts of CpG sites. This aims to make a direct impact in the future of cancer care and analyzing patterns in data related to methylation levels.

*(continue to page 44)*

*(continued from page 43)*

Wording for Poster (written but not used for abstract):
Molecular biology methods heavily rely on controls in order to fully understand, analyze and interpret results. Widely used applications dependent on these controls range from next generation sequencing (NGS) and quantitative PCR (qPCR) to enzyme engineering and genome editing. Another emerging application dependent on controls is determining methylation patterns found in cancer cells. DNA methylation modifications on specific nucleotides can affect gene expression levels through epigenetic processes, demonstrating changes in levels being linked to specific cancer types. Cytosine methylation is most commonly found at CG sequences in the genome, referred to as a CpG site, and is widely used to regulate gene expression in a cell-type specific manner. Defining the level of expression in a normal cell versus a cancer cell by using a control sequence at a specific level can truly be life altering for patients.
https://ascopubs.org/doi/10.1200/JCO.2020.38.15_suppl.3052

# PacBio cDNA concatenation to increase sequencing throughput for RNA isoform discovery



**Anthony R. Miller[1], Saranga Wijeratne[1], Maria Elena Hernandez Gonzalez[1], Kelli Roach[1], James R. Fitch[1], Benjamin J. Kelly[1], Peter White[1,2], Catherine E. Cottrell[1,2,4], Elaine R. Mardis[1,2,3], Richard K. Wilson[1,2]**

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
4 Department of Pathology, The Ohio State University College of Medicine, Columbus, OH, USA

The mutational spectrum associated with oncologic diseases encompasses single-nucleotide variation, insertion-deletion events, and structural variation. Additionally, tumorigenesis can result in RNA dysregulation, including alternative splicing of transcripts. Pacific Biosciences (PacBio) long-read RNA-Seq (Iso-Seq) protocol can help resolve such transcriptomic events. This approach obviates the need for RNA fragmentation, resulting in the preservation of expressed exonic order of transcripts, and provides full-length isoform characterization to help elucidate novel disease-associated isoforms.

A challenge of using Iso-Seq for discovery of disease-associated isoforms is limited throughput. As a result, detection of moderate to lowly expressed isoforms can be more difficult compared to short-read RNA-Seq methods. The recent introduction of MAS-ISO-seq (Al'Khafaji et al.), a technique for concatenating single molecules for long-read sequencing, provides a unique opportunity to boost throughput. We have modified the MAS-ISO-seq protocol to support bulk Iso-Seq concatenation.

*(continue to page 46)*

*(continued from page 45)*

In our approach, we first size-select cDNA molecules into two separate fractions, molecules greater than 2,000 bp and molecules between 500 to 2,000 bp. The cDNA molecules greater than 2,000 bp are used for dimer concatenation and the remaining 500 to 2,000 bp molecules are used for generating a tetramer molecule. Post concatenation, the two pools are combined after PacBio SMRTbell prep and sequenced on a SQIIe 8M SMRTcell.

To evaluate our method, we prepared a concatemer and non-concatemer control sample comprised of commercial reference RNA (Spike-In RNA Variants; SIRV-4) with known isoform complexity mixed with Human Brain Reference RNA (HBR_SIRV). After sequencing and deconcatentation using Skera, we generated 8,931,803 individual HiFi reads from our HBR_SIRV concatemer sample, a 4.5-fold increase in throughput over the control (1,965,948 HiFi reads). Of the 8,931,803 HiFi reads, 2,066,749 were generated from the dimer (average cDNA size of 2,215 bp) and 6,865,054 from the tetramer (average cDNA size of 1,085 bp). In total, we detected a 16-fold increase in the SIRV isoform representation from the concatemer sample compared to the non-concatemer control, respectively 294,630 and 18,091 HiFi reads.

Here, we describe a methodology for increasing the throughput of bulk Iso-Seq through an adaptation of the MAS-ISO-seq protocol to help identify novel disease associated isoforms that lack abundant expression.

Al'Khafaji, A.M., J.T. Smith, K.V. Garimella, M. Babadi, M. Sade-Feldman, M. Gatzen, S. Sarkizova, M.A. Schwartz, V. Popic, E.M. Blaum, A. Day, M. Costello, T. Bowers, S. Gabriel, E. Banks, A.A. Philippakis, G.M. Boland, P.C. Blainey, and N. Hacohen: High-throughput RNA isoform sequencing using programmable cDNA concatenation. bioRxiv 2021:2021.10.01.462818.

# Polyethnic-1000: A New York–based initiative to advance cancer genomics through recruitment of diverse racial and ethnic populations



**Nicolas Robine[1], Lara Winterkorn[1], Onyinye Balogun[1,3], Melissa Davis[1,3], Michelle Mehallow[1], Tim Chu[1], Charles Sawyers[1,4], David Tuveson[1,2], Harold Varmus[1,3]**

1 New York Genome Center;
2 Cold Spring Harbor Laboratory,
3 Weill Cornell Medicine;
4 Memorial Sloan Kettering Cancer Center

Recent advances in DNA sequencing technologies have revolutionized approaches to the prevention, risk assessment, early detection, diagnosis, and treatment of cancers. However, our current knowledge about tumor biology, cancer risk, and response to treatment is limited due to significant underrepresentation of non-European populations in genomic cancer research, including clinical trials. The vast majority of samples in publicly available genomic databases have been derived from patients of European descent. These inequities limit our understanding of cancer and the impact of ancestry on the various manifestations of this disease. Moreover, exclusion of minoritized populations may exacerbate health disparities and stymie their ability to equally benefit from participation in trials or the innovations that result from such studies.

Through the Polyethnic-1000 (P1000) initiative, we seek to leverage the racial and ethnic diversity within New York City to conduct initial investigations that address the existing disparities in cancer genomics studies. We are performing tumor-normal whole genome sequencing and tumor RNA-Seq on 1000 cases of bladder, breast, colon, endometrial, lung, pancreatic, and prostate cancers, as well as multiple myeloma. Although patients are enrolled based on self-reported race and ethnicity, we will estimate genetic ancestry from their normal genomes. Analyses will be performed with NYGC's germline and somatic pipelines, and data will be compared to publicly accessible cohorts.

The resulting data set will be among the largest and most diverse collection of clinically annotated full-genome pairs available in oncology. It will be housed at the NYGC, and rapidly shared with the entire research community in protected-access repositories. We expect sample collection to be completed in early 2023 and some of the cancer-specific cohort analyses to be well advanced by the time of the meeting.

*(continue to page 48)*

*(continue from page 47)*

P1000 has established a framework to enhance interactions among our region's academic and health centers to advance cancer genomics. These efforts should improve and widen the use of genomics for all, especially currently underserved racial and ethnic populations.

# Process description and quality control gates involved in adapting a standard WGS service to a clinically accredited automated workflow, and preliminary results on clinical validation of automated plasma WGS workflow



**Lubaina A. Kothari[1], Savo Lazic[1], Theodore Chan[1], Sharanjit Singh[1], Andrea Bevan[1], Faridah Mbabaali[1], Madhuran Thiagarajah[1], Bernard Lam**

1 Ontario Institute for Cancer Research, Toronto, ON, Canada

In recent years, rapid advancements in precision oncology research have paved the way to discoveries of novel biomarkers and enhanced knowledge of early cancer diagnosis and monitoring. These developments have accelerated the need to develop detection assays to harness these ever-evolving findings. The Translational Genomics Laboratory (TGL), part of the Ontario Institute of Cancer research (OICR), a state-of-the-art not-for-profit Canadian Public Translational Cancer Centre, is dedicated to leading the way in clinical cancer care. With our R&D arm, we focus our efforts towards onboarding the latest oncogenomic assays, standardizing these protocols for scale, and integrating data analysis to our custom informatics pipeline. Our clinical arm is committed to rapidly transitioning these developed assays in compliance with CAP/CLIA accreditation standards to deliver reliable testing to clinicians worldwide. In this presentation, we will cover the entire process taken to adapt our standardized Whole Genome Sequencing (WGS) service to a clinically accredited automated workflow on the Sciclone G3 NGSx Workstation. This includes receipt/inspection, extraction, library preparation, library qualification, MiSeq and NovaSeq QC, informatics pipeline with variant interpretation and the final report. Our WGS is optimized to handle a variety of sample types like fresh frozen (FF) and formalin-fixed paraffin-embedded (FFPE). Recent studies have also highlighted a shift from solid tumors to liquid biopsies for non-invasive cancer monitoring in initial stages, assessing of treatment options and tracking recovery and relapse. We will also present our current work on adapting this process towards a clinically validated and automated noninvasive liquid biopsy sequencing service for high sensitivity cancer monitoring using a cell-free genome-wide sequencing approach at standard sequencing depths.

# Single-cell whole transcriptome profiling of human FFPE tumor blocks with Chromium Fixed RNA Profiling



**Jawad Abousoud**

10x Genomics

Archived formalin-fixed, paraffin-embedded (FFPE) tissue samples are the largest repository of human tumor specimens in academic and clinic settings. Gene expression profiling of these FFPE samples can provide critical insights into disease etiology. However, existing technologies are either incompatible with FFPE tissues due to RNA degradation or there is a need to improve data resolution to single-cell level.

The new Chromium Fixed RNA Profiling assay now enables single cell FFPE sequencing (scFFPE-Seq) for detection of whole transcriptome gene expression in FFPE tissues via in situ hybridization of probes to the RNA transcripts. Hybridized probes are ligated, barcoded, and sequenced, which allows highly sensitive detection of transcripts even in FFPE preserved tissues with lower sample quality compared to fresh or flash frozen tissues. To evaluate this technology, we dissociated FFPE sections from 35 human blocks, including lung, breast, prostate, ovary, colorectal, glioblastoma, and melanoma cancer samples, and analyzed them using the Chromium Fixed RNA Profiling assay. The data showed distinct cell clustering and identified representative cell types in the FFPE samples, comparable to published single cell data for each tissue type. Additionally, the data quality was consistent across different sections of the same FFPE block, indicating the robustness and reliability of the assay.

To provide an additional layer of biological insight from the spatial perspective, we performed Visium CytAssist Spatial Gene Expression for FFPE using sections from the same FFPE blocks. We were able to detect comparable levels of marker gene expression after integrating the spatial data with scFFPE data, and found that the same cell types and proportions were identified in both data sets.

In summary, we demonstrated the power of the Chromium Fixed RNA Profiling assay in supporting scFFPE-Seq to unlock the biology preserved in archived FFPE tumor samples. The assay truly expands the capabilities of 10x Genomics' Chromium platform, enabling cross-assay compatibility with Visium CytAssist for FFPE samples and serves as a powerful tool to facilitate discoveries in disease progression and therapeutic target development.

# Single-cell resolved spatial architecture of non–small cell lung cancer in 3D



**Tancredi Massimo Pentimalli[1], Simon Schallenberg[2], Ilan Theurillat[1], Gwendolin Thomas[1], Anastasiya Boltengagen[1], Youngmi Kim[3], Sean Kim[3], Sarah Murphy[3], Mark Gregory[3], Nikos Karaiskos[1], Ivano Legnini[1], Stefano Piccolo[4,5], Andrew Woehler[6], Frederick Klauschen[2], Nikolaus Rajewsky[1]**

1 Laboratory for Systems Biology of Regulatory Elements, Berlin Institute for Medical Systems Biology (BIMSB), Max Delbruck Center for Molecular Medicine in the Helmholtz Association (MDC), 10115, Berlin, Germany
2 Institute of Pathology, Charité-Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität Zu Berlin and Berlin Institute of Health, 10117, Berlin, Germany
3 NanoString® Technologies, Seattle, WA 98109, USA
4 Department of Molecular Medicine, University of Padua, Padua, Italy
5 IFOM ETS, the AIRC Institute of Molecular Oncology, Milan, Italy
6 Systems Biology Imaging Platform, Berlin Institute for Medical Systems Biology (BIMSB), Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), 10115

Immune checkpoint inhibitors (ICI) have revolutionized the treatment of non-small cell lung carcinoma (NSCLC), however response rates remain unsatisfactory and predicting the patients who will benefit from treatments is challenging. In this regard, deeper understanding of the tumor microenvironment (TME) is essential to identify patient-specific immunotherapeutic targets. To understand the cellular composition within the TME, three-dimensional (3D) spatial organization and cell-cell communication in one aggressive NSCLC tumor, we employed the CosMx™ Spatial Molecular Imager to quantify and localize transcripts of 1,000 target genes in 6 consecutive FFPE sections.

The high-quality data and the single molecule resolution enabled the assignment of more than 100 million transcripts in 350,000 segmented cells. We then computationally aligned the 2D slices and generated a 3D high-resolution molecular map of the tumor. This allowed us to analyze TME heterogeneity in 3 dimensions. Analyses of 3D cellular neighborhoods revealed 7 major multicellular niches characterized by specific combinations of cell types. The examples of niches include the tumor bed, macrophage islands, tumor-associated stroma and dendritic cell hubs.

*(continue to page 52)*

*(continued from page 51)*

We further leveraged the 3D niches to functionally link transcriptomic states, cellular neighborhoods and cellular morphology. For example, by comparing tumor cells in the tumor bed with tumor cells located in the stroma, we identified an EMT phenotype in infiltrating tumor cells. Furthermore, we show that fibroblasts in the tumor-associated stroma upregulate collagen production, increasing in size, and contribute to tumor growth. Cytotoxic T cells (CTL) were highly abundant in the sample but were mainly restricted to dendritic hubs colocalizing with T regulatory cells and dendritic cells.

Spatial analysis of receptor-ligand interaction identified multiple immune inhibitory signals, including canonical PD-L1/PD-1 signalling but also non-canonical immune suppressive Galectin 9/Tim-3 and MIF/CD44 interactions. This could explain why, despite their high numbers, CTL failed in eliminating tumor cells. Our data suggest that anti-PD-1 with anti-Tim3 combination therapy , which currently evaluated in clinical trials, might have been beneficial for the patient.
In summary, our data suggest that single-cell resolved, 3D molecular pathology will help in the identification of personalized immunotherapeutic targets in cancer patients.

# Single-molecule epigenetic profiling of early-stage kidney cancer



**Prof. Yuval Ebenstein**[1]

1 Tel-Aviv University, Department of Chemistry, Tel-Aviv, Israel

Cancer genomes often harbor multiple DNA rearrangements that short-read sequencing is blind to. Long reads generated by optical genome mapping provide a unique opportunity to study long-range rearrangements common in cancer, such as medium to large structural variations, copy number variations, and complex fusions, as well as simultaneous alterations in epigenetic signatures in genes and their distal enhancers in single-cell level. This is the first time a tumor and a normal adjacent tissue are mapped using Bionano-Genomics technology to locate somatic genetic and epigenetic aberrations. This method can be applied to a wide range of solid tumors and aid in unraveling some of the complexity of the cancer genome and epigenome.

Clear cell renal cell carcinoma (ccRCC) is the most common subtype of kidney cancer. Most ccRCC cases demonstrate distinctive changes to the short arm of chromosome 3, most often involving the Von Hippel–Lindau (VHL) tumor suppressor gene. In order to identify disease-related alterations that appear at an early stage, we performed a comprehensive genetic, cytogenetic, and epigenetic analysis in a ccRCC tumor and adjacent normal tissue from a patient with early-stage ccRCC. We used exome sequencing and optical genome mapping to identify genetic aberrations as single nucleotide polymorphism, structural and copy number variations in the tumor and in the adjacent normal tissue. Then, we mapped the single-molecule methylome and hydroxymethylome of the two samples and used these epigenetic maps to locate differentially modified regions along the genome. A driver somatic variant was discovered at the VHL gene in the tumor sample. Additionally, the methylation and hyroxymethylation profiles revealed multiple differential regions, some located in genes, promoters and distal enhancers of genes known to be associated with ccRCC. Metabolic pathways were significantly enriched among differentially modified gene bodies and enhancers. Moreover, significant global epigenetic differences were detected between the two samples, and a correlation between epigenetic signals and gene expression was found.

# Spatial transcriptomic profiling of prostate cancer reveals zone-specific androgen receptor signaling and tumor immune microenvironment



**Yan Liang[1], Andy Nam[1], Parth Patel*[2], Srinivas Nallandhighal*[2], David Scoville[1], Lynn Tran[3], Brittney Cotta[2], Aaron M. Udager[4,5,6], Arvind Rao[6,7], Ganesh S. Palapattu[2,6], Vipulkumar Dadhania[4], Sethu Pitchiaya[2,4,5], Simpa S. Salami[2,5,6]**

1 NanoString® Technologies, Inc., Seattle, WA, USA.
2 Department of Urology, Michigan Medicine, Ann Arbor, Michigan, USA.
3 Medical College of Georgia, Augusta, Georgia, USA
4 Department of Pathology, Michigan Medicine, Ann Arbor, Michigan, USA.
5 Michigan Center for Translational Pathology, Michigan Medicine, Ann Arbor, USA
6 University of Michigan Rogel Cancer Center, Ann Arbor, Michigan, USA.
7 Department of Computational Biology, Michigan Medicine, Ann Arbor, Michigan, USA.
*Equal contribution

**Background:** Transitional zone prostate cancer (PCa) accounts for approximately 30% of disease, tends to have higher PSA and larger size, but lower odds of seminal vesicle invasion, extra-capsular extension, recurrence rates compared with peripheral zone (PZ). Underlying mechanisms for these differences are poorly understood. We performed spatial transcriptomic profiling to elucidate the biology of TZ and PZ PCa.

**Methods:** With samples from three PCa patients after radical prostatectomy (one each with PZ, TZ, and both PZ and TZ tumors), we used NanoString's Whole Transcriptome Atlas on the GeoMx® Digital Spatial Profiler (DSP) to assess gene expression in multiple regions of interest (ROI) per patient (42 total tumor ROIs). Morphology markers SYTO13 (nucleus), PanCK (epithelium), SMA (stroma), and CD45 (immune cells) were used to select ROIs. Raw counts for 17,128 genes were imported to R for downstream data analyses. Counts were Q3 normalized and scaled. We performed differential expression analysis using a linear model fit by empirical Bayes moderation, gene set enrichment analysis (GSEA) by cancer hallmarks, XCell gene sets for pathway enrichment and immune cell deconvolution using CIBERSORT.

*(continue to page 55)*

*(continue from page 54)*

**Results:** There were grade group (GG) 4 (n=10) and 5 (n=10) tumors in PZ and GG 3 (n=10), 4 (n=11) and 5 (n=1) in TZ . We observed distinct gene expression profile between PZ (n = 20) and TZ (n=22) tumors. Androgen receptor (AR) signaling was significantly higher in TZ PCa ROIs compared to PZ in both GSEA (FDR < 5%) and the androgen subcomponent of the genomic prostate score ($p<0.001$), regardless of grade or cellular composition. CIBERSORT's absolute immune signature scores were computed for GG4 tumors and were significantly higher in PZ GG4 tumors compared to TZ GG4 tumors. Notably, CD4+ memory T cells were significantly higher in PZ GG4 regions compared to TZ GG4 tumor regions ($p<0.05$).

**Conclusions:** We identified higher AR signaling in TZ tumors, rationalizing historically observed higher PSA levels compared to PZ. We also observed increased immune infiltration within PZ microenvironment compared to TZ. Further studies will identify robust biomarkers to distinguish PZ from TZ and shed light on PCa immunomodulation, with potential immunotherapy implications.

**Disclosures**

AR consults for Genophyll, LLC and serves as member of Voxel Analytics LLC.
AN, DS, and YL work for NanoString® Technologies

# Spatially resolved heterogeneity of the consensus molecular subtypes of colorectal cancer

**Nadine Kumpesa[1], Bettina Amberg[1,2], Marc Sultan[1], Alberto Valdeolivas[1*], Kerstin Hahn[1*]**

1 Roche Pharma Research and Early Development, Roche Innovation Center Basel, Basel, Switzerland
2 Institute of Animal Pathology and Bern Center for Precision Medicine Vetsuisse Faculty, University of Bern, Switzerland
* These authors contributed equally

Colorectal cancer (CRC) is a leading cause of cancer-related death worldwide. CRC shows profound heterogeneity which leads to differential treatment responses among patients, thus pressing the need to refine the current classification and develop more targeted therapies. Recently, several studies addressed the heterogeneity of this disease, resulting in the widely accepted classification into four consensus molecular subtypes (CMS).

To further study the heterogeneity of the CMS, we here collected 14 samples from a diverse cohort of seven CRC patients and applied Spatial Transcriptomics (ST). We integrated our ST data with a recently published single-cell RNA sequencing (scRNA-seq) CRC dataset using a deconvolution-based approach.

Our results provide novel insights into the spatial composition and heterogeneity of CRC and its CMS. We identified that CMS1 and CMS2 were distinctly confined to the neoplastic areas, whereas CMS3 signatures were exclusively associated with the non-neoplastic mucosa. In addition, the pro-inflammatory landscape of CMS1 in comparison with the poor immune infiltration in CMS2 was delineated. We demonstrated the power of ST to spatially depict the heterogeneity of CRC tumors. For instance, we highlighted the presence of intra-tumor CMS mixed phenotypes in different anatomical locations and linked them to spatially variable molecular features, such as pathway activity. Our results describe common cell-to-cell communication events between CMS2 tumors and their neighborhood, and inferred signaling pathways and transcriptional programs modulating tumor progression.

*(continue to page 57)*

*(continued from page 56)*

Our approach provides new insights on CRC spatial heterogeneity and therefore conveys new opportunities for the development of more tailored treatments.

**Conclusions:** We identified higher AR signaling in TZ tumors, rationalizing historically observed higher PSA levels compared to PZ. We also observed increased immune infiltration within PZ microenvironment compared to TZ. Further studies will identify robust biomarkers to distinguish PZ from TZ and shed light on PCa immunomodulation, with potential immunotherapy implications.

**Disclosures**
AR consults for Genophyll, LLC and serves as member of Voxel Analytics LLC.

# Ultra–high-plex spatial proteogenomic investigation of giant-cell glioblastoma multiforme immune infiltrates reveals distinct protein and RNA expression profiles



**Alyssa Rosenbloom[1], Shilah A. Bonnett[1], Giang Ong[1], Mark Conner[1], Aric Rininger[1], Daniel Newhouse[1], Felicia New[1], Hiromi Sato[1], Chi Phan[1], Saskia Ilcisin1, John Lyssand[1], Gary Geiss[1], and Joseph M. Beechem[1]***

1 NanoString® Technologies, Seattle, WA, USA

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. Most spatial technologies yield single analyte proteomic or transcriptomic datasets from separate formalin-fixed paraffin-embedded (FFPE) tissues sections. Multiple studies have demonstrated a poor correlation between RNA expression and protein abundance owing to transcriptional and translational regulation, target turnover, and post-translational protein modifications. Therefore, a workflow that accurately measures RNA and protein simultaneously within a single tissue section with distinct spatial context is critical to a more complete biological understanding of cellular interactions and activities. Such multimodal omic datasets of protein and DNA or RNA have been termed "spatial proteogenomics".

Here we present a novel spatial proteogenomic (SPG) assay on the GeoMx® Digital Spatial Profiler platform with NGS readout that enables ultra high-plex digital quantitation of proteins (147-plex) and RNA (whole transcriptome, > 18,000-plex) from a single FFPE sample. We demonstrated high concordance, R > 0.85, and minor change in sensitivity (<11%) between the SPG assay and the single analyte GeoMx Whole Transcriptome Atlas and GeoMx NGS Protein assays. We used the SPG assay to interrogate 23 different glioblastoma multiforme samples across 4 pathologies. We observed clustering of both RNA and protein based on cancer pathology and anatomic location. The in-depth investigation of giant cell glioblastoma multiforme (gcGBM) revealed distinct protein and RNA expression profiles compared to that of glioblastoma multiforme (GBM). Spatial proteogenomics allowed simultaneous interrogation of critical protein post-translational modifications alongside whole transcriptomic profiles within the same distinct cellular neighborhoods.

*(continue to page 59)*

*(continue from page 58)*

Within our dataset, we observed >2-fold higher protein expression levels of phospho-GSK3β (Ser9) in gcGBM compared to GBM. Inactivation of GSK3β through phosphorylation has been shown to enhance proliferation of GBM cells. We also observed differential protein expression phosphorylated Tau variants. Phospho-Thr231 Tau was >2-fold higher in GBM compared to gcGBM. Associated with neurodegenerative Alzheimer's disease, changes in Tau expression and phosphorylation have also been observed in glioblastoma. Our study exemplifies the utility of the SPG assay in expanding our understanding of glioblastoma multiforme molecular pathology.

# Use of synthetic CNV fragments to mimic copy number alterations for ctDNA reference standards



**Jason Corwin[1], Patrick Cherry[2], Shawn Gorda[3], Michael Bocek[1], Jean Challacombe[1], Derek Murphy[2], Esteban Toro[2]**

1 Twist Bioscience, Research & Development Bioinformatics, South San Francisco, California, USA
2 Twist Bioscience, Research & Development, South San Francisco, California, USA
3 Twist Bioscience, Office of Product Management, South San Francisco, California, USA

Diagnostic assays of liquid biopsies hold great promise for cancer detection and monitoring. To aid the advancement of this technology, Twist Bioscience previously released the Pan-cancer Reference Standard, a cell free DNA (cfDNA) mimic reference material with varying allelic fractions of synthetically-printed single nucleotide variations (SNVs), insertion-deletions (INDELs), and structural variants (SVs). However, many cancers are characterized by changes in the copy number of genes that support growth and survival of cancer tissue—dubbed copy number variants, or CNVs. To aid assay development for detection of CNVs, we designed, synthesized, and characterized an additional control supplemented into the Pan-cancer Reference Standard: a CNV for the exonic regions of the ERBB2 gene (also known as HER2). ERBB2 is often amplified and overexpressed in invasive carcinomas, including breast, ovarian, stomach, bladder, salivary, and lung cancers, making it a desirable variant to include in reference materials.

Similar to the design of the synthetic variants in the Twist Pan-cancer Reference Standard, the ERBB2 CNV DNA is developed to closely mimic the size and content of native ctDNA while providing a wide selection of common and rare cancer targets for analytical validation. It mimics the length distribution of cfDNA, with a peak at 167 bp ± 5 bp for diversity. The synthetic CNV DNA tiles over the exonic intervals, for uniform gene coverage. Variant fragments were printed, quantified, and uniformly pooled followed by spiking into DNA derived from a single, highly NGS-characterized donor as background. The final admixture of ERBB2 is ddPCR-verified for copy number amplification.

The Twist Pan Cancer Reference Standard continues to provide a valuable highly-multiplexed solution to assay developers seeking reference materials for a wide array of cancer variants, and is now further improved with the addition of copy number variants.

# Large-scale and high-throughput spatial omics to deeply characterize cancer and normal tissue



**Joachim H. Schmid[1], Sarah E Church[1], Margaret Hoang[1], Gokhan Demirkan[2], Kyla Teplitz[2], Joseph M Beechem[1]**

1 NanoString® Technologies Inc, Research & Development, Seattle, WA, USA
2 NanoString Technologies Inc, Marketing, Seattle, WA, USA

To understand heterogeneity of human tissue and identify biomarkers for predicting response to therapies it is necessary to evaluate thousands of patient samples. Here we present a process for high-throughput spatial analysis of the whole transcriptome and over 100 proteins for thousands of clinically relevant cancer and normal samples using the GeoMx® Digital Spatial Profiling (DSP) platform. From these analyses we are able to characterize the tumor microenvironment in separate compartments of the interepithelial, stromal and immune cell rich areas of multiple solid tumor indications and subtypes. Similarly, in normal tissue we are able to deeply characterize tissue specific histology such as glomeruli in kidney, hippocampus in brain or peyer's patches in colon tissue.

Here we present a method to spatially analyze 60 FFPE samples on a DSP instrument per week using established automated procedures for each step of the DSP workflow. This process incorporates region of interest (ROI) and area of illumination (AOI) selection, slide processing, the DSP instrument run, sequencing preparation, sequencing and data analysis. Pre-selection of up to 12 ROIs/AOIs per sample is performed on sequential H&E slides, which then can be rapidly overlayed onto the DSP immunofluorescent scans, easily placed and automatically segmented on specific tissue such at cytokeratin-positive tumor and cytokeratin-negative stroma. Slide processing is done automatically using the Leica BOND system for 20 slides with 3 samples per slide. All annotations chain of custody can be tracked during the process using a barcode reader. After DSP on-instrument processing, collection plates can then be stored until ready for automated pre-sequencing processing. Following sequencing, data is processed and uploaded to AtoMx™ Spatial Informatics Platform for data QC, analysis, collaboration and storage.

*(continue to page 62)*

*(continued from page 61)*

This high throughput workflow spatially analyzing both protein and whole transcriptome gene expression allows for the creations of databases and consortiums containing meta-data for large numbers of tissues. These databases can be used to train new AI methods, to deeply evaluate tissue neighborhoods, and in the case of clinical trials, to discover biomarkers to understand mechanism of disease and potentially help improve selection of therapies in the clinic.

# High-speed multiomic spatial phenotyping of immunotherapy responses in head and neck cancer



**Oliver Braubach[1], James Monkman[2], Ken O'Byrne[3], Brett Hughes[4], Niyati Jahveri[1], Ning Ma[1], Bassem Ben Cheikh[1], Arutha Kulasinghe[2]**

1 Akoya Biosciences, Massachusetts, MA, USA
2 The University of Queensland, Brisbane, Queensland, Australia
3 The Princess Alexandra Hospital, Brisbane, Queensland, Australia
4 The Royal Brisbane and Women's Hospital, Brisbane, Queensland, Australia

**Background:** Immune check point inhibitors (ICI) have proven to be game-changing treatments for mucosal head and neck squamous cell cancer (HNSCC). Emerging successes with anti-PD-1/PD-L1 ICI therapy have led to durable responses and prolonged survival in human papillomavirus-positive (HPV+) and negative (HPV-) patients. There is now an urgent need for predictive biomarkers to guide patient selection for highly targeted ICI therapies as currently available diagnostic biomarkers have limited value. The tumor microenvironment (TME) composition, contexture, and cellular architecture are now recognized as key to understanding immune responsive and resistant phenotypes.

**Methods:** In this study, we used single-cell, multiomic spatial phenotyping to characterise the TME using unbiased whole slide imaging of metastatic/recurrent HNSCC tumors from a cohort of n=40 patients treated with Pembrolizumab / Nivolumab. The discovery cohort consisted of patients who had complete vs. partial vs. stable vs. progressive responses to ICI therapy. We first analyzed tissues using an ultrahigh-plex antibody panel (60+antibodies), imaged with the PhenocylcerFusion platform (Akoya Biosciences). We then conducted additional whole-slide spatial biology experiments with a mixed protein / RNA detection panel to obtain multiomic spatial signatures that could offer cues as to which treatment matches best with certain outcome groups.

*(continue to page 64)*

*(continued from page 63)*

**Results:** Our study identified stromal, immune, and metabolic tissue signatures associated with resistance to immunotherapy. Most notably, high expression of VISTA within the stromal compartment was associated with poor overall survival. Moreover, spatial metabolic profiling identified defined areas of high and low metabolic activity reflecting the immune-cell landscapes associated with therapy resistance.

**Conclusions:** Our study demonstrates the power of unbiased multiomic spatial phenotyping with whole-slide imaging to identify biomarkers associated with response to ICI therapy in HNSCC.

# Orthologs of targetable human cancer somatic variants are detectable in canine cell-free DNA

**Ilya Chorny[1], Kristina M. Kruglyak[1], John A. Tynan[1], Susan Hicks[1], Jill M. Rafalko[1], Todd A. Cohen[1], Allison L. O'Kell[1], Jason Chibuk[1], Angela L. McCleary-Wheeler[1], Andi Flory[1], Daniel S. Grosu[1], Dana W.Y. Tsui[1]**

1 PetDx, La Jolla, CA

Genotype-matched therapeutics are widely used in the treatment of human cancers. OncoKB (Oncology Knowledge Base) is the first FDA-recognized database that contains information about these targetable somatic variants and corresponding therapeutic agents, including levels of evidence. In human cancer patients, the variants targeted by these agents can be detected in cell-free DNA through liquid biopsy testing. Many of the human variants found in the OncoKB database have orthologs in the canine genome. This study was conducted to determine the feasibility of detecting these orthologs in blood of cancer-diagnosed dogs.

A representative cohort of >600 client-owned dogs with a variety of cancer diagnoses across all disease stages had pre-treatment plasma samples collected at the time of enrollment in a clinical validation study for a canine liquid biopsy test. A subset of dogs also had matched tumor tissue available for analysis.

Blood samples (and tumor samples, when available) were subjected to DNA extraction, proprietary library preparation, and next-generation sequencing. Sequencing data were analyzed using an internally developed bioinformatics pipeline to detect genomic alterations associated with the presence of cancer. Only canine orthologs of variants in the OncoKB database were evaluated in this study. Variants were called if they were observed in plasma at an allele frequency of at least 0.5%, with at least 6 unique molecules. Variants identified from plasma were subsequently genotyped in tissue without additional variant-level filtering.

Analysis of pre-treatment plasma identified canine orthologs of OncoKB variants in ~8% of subjects, and in 15% of these, the dog had a cancer type that matched the human indication for the detected variant. Of dogs in which an ortholog was identified in plasma, 44% had matched tumor tissue; and the variant observed in plasma was confirmed in tissue in >50% of these subjects.

*(continue to page 66)*

*(continued from page 65)*

This study demonstrates the feasibility of detecting canine orthologs of targetable human cancer somatic variants in the blood of cancer-diagnosed dogs using next-generation sequencing. Though it is currently unclear how human databases can be used to inform targeted treatment decisions in dogs, these findings may have relevance for comparative oncology studies and future precision therapeutics development for both species.

# Genetic ancestry correlates of the somatic mutational landscape from tumor profiling data of 100,000 cancer patients



**Francisco M. De La Vega, Brooke Rhead, Yannick Pouliot, and Justin Guinney**

Tempus Labs, Inc., Chicago, IL, USA

Cancer genomic studies have underrepresented minorities and individuals of non-European descent, thus limiting a comprehensive understanding of disparities in the diagnosis, prognosis, and treatment of cancer among these populations. Furthermore, the social constructs of race and ethnicity are far from precise categories to understand the biological underpinnings of such differences. In this study, we use a large real-world data (RWD) patient cohort to examine associations of genetic ancestry with somatic alterations in cancer driver genes. We inferred genetic ancestry from approximately 100,000 de-identified records from cancer patients of diverse histology who underwent tumor genomic profiling with the Tempus xT 648-gene next-generation sequencing (NGS) assay. We used 654 ancestry informative markers selected to overlap the capture regions of the assay to infer global ancestry proportions at the continental level: Africa (AFR), America (AMR), Europe (EUR), East Asia (EAS), and South Asia (SAS). While most patients are of European descent (72%), our cohort includes 4.7 and 3.8-fold more patients with substantial (>50%) AFR and AMR ancestry, correspondingly, than TCGA. Logistic regression was used to directly test for associations between continental ancestry proportions and presence of nonsynonymous somatic mutations and copy number alterations (CNA) in cancer genes, controlling for assay version, gender and age. P-values were adjusted for multiple testing by the Benjamini-Hochberg method to control the false discovery rate at 5%. We identify 7 significant associations with small somatic mutations and 15 with CNAs with non-European ancestries (per 20% increase in each ancestry proportion; all $p<0.0001$). Among others, we found associations between small somatic mutations in CTNNB1 with EAS ancestry (OR=1.44); EGFR with EAS (OR=1.49), and AMR (OR=1.78) ancestries in lung cancer, and ASXL1 with AMR ancestry in brain cancer (OR=2.48). Furthermore, we identified several associations between ancestry and CNAs in MTAP with AMR (OR=1.45) and EGFR and SAS (OR=1.46) in lung cancer, among others.

*(continue to page 68)*

*(continued from page 67)*

Finally, we observed a reduction in actionable mutations (OncoKB levels 1,2, and R1) with AFR ancestry in BRAF (OR=0.73) and EGFR (0.77) in colorectal and lung cancers, correspondingly. Our results support the use of genetic ancestry inference on RWD to improve upon the use of race and ethnicity to understand the impact of ancestry into mutational processes that influence cancer incidence, progression, and outcomes.

# Functional drug testing of pediatric acute lymphocytic leukemia cells with single-cell transcriptomic readout maps cellular drug response

**Jessica Nordlund[1,2], Henrik Gezelius[1,2], Anna Pia Enblad[1,3], Anders Lundmark[1,2], Martin Åberg[1,4], Kristin Blom[1,4], Jakob Rudfeldt[1,4], Arja Harila-Saari[3], Amanda Raine[1,2], Claes Andersson[1,4]**

1 Department of Medical Sciences Uppsala University, Uppsala, Sweden
2 Science for Life Laboratory, Uppsala University, Uppsala Sweden
3 Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden
4 Department of Clinical Chemistry and Pharmacology, Uppsala University Hospital, Uppsala, Sweden

Functional precision medicine involves ex vivo measurement of drug response in order to tailor treatments for individual patients. Here, we functionally annotated drug resistance using the glucocorticoid-resistant E/R+ REH acute lymphoblastic leukemia (ALL) cell line as a cellular model in response to 25 drugs. We evaluated three different technologies for barcoding selected drug conditions with a single-cell gene expression (scRNA-seq) readout. We sequenced >30,000 single cells with three current scRNA-seq methods and benchmarked the technologies against each other as well as to bulk RNA-seq. Each of the scRNA-seq approaches accurately reflected gene expression changes in the system, with high cell recovery and reproducible tagging of the different drug conditions. However, differences in sensitivity and implementation feasibility were discerned. Furthermore, we identified a substantial and reproducible transcriptional response to fludarabine at a clinically relevant concentration (0.56 µM). Fludarabine is a drug of particular interest for treatment of high-risk ALL.

# A spatial atlas of triple-negative breast cancer in women of African ancestry



**Jasmine Plummer[1], Omotoso Ayodele[2], Nadezhda Nikulina[3], Felipe Dezem-Segato[1], Maycon Marcao[1], Oliver Braubach[3], AC3 4, Matthew Schlumbrecht[2], Sophia George[2].**

1 Center for Spatial OMICS, St Jude Children's Research Hospital, Memphis TN
2 Sylvester Comprehensive Cancer Center, University of Miami Health Systems, Miami FL
3 Akoya Biosciences, The Spatial Biology Company, Marlborough, Massachusetts, USA
4 African Caribbean Cancer Consortium

The US Black population is composed of both US-born Black and immigrant Black populations from the Caribbean and Africa. Normal tissues in Black individuals independent of their country of birth or residence are woefully understudied. Black individuals disproportionately develop aggressive pathologic diseases, which are treatment refractory or resistant, leading to premature deaths. In women, breast cancer is more common among US Black women, as well as the most common non-viral driven cancer in African and Caribbean women. Furthermore, black women develop this disease younger than other ancestral groups and have a higher incidence of metaplastic and triple-negative breast cancer – aggressive pathologies. With the African-Caribbean Cancer Consortium (AC3) and Transatlantic Gynecologic Cancer Research Consortium, we have created a multiomic spatial atlas of triple-negative and other breast cancers across Africa, the Caribbean, and among US-Black individuals.

To this end, we performed ultrahigh-plex RNA and protein spatial phenotyping on the PhenoCycler-Fusion (PCF). The PCF is a fast end-to-end spatial biology platform that enables whole-slide spatial readouts of RNA and protein moieties at single-cell resolution. Multiomic spatial phenotyping of tissues allowed for the detection of novel cell populations associated with a given African ancestry linking unique immune/stromal cell types to the outcome and severity of breast cancer

With this study, we aim to develop a benchmark that confidently measures and interprets ancestral genomic differences at the cellular level. Deciphering the relationship between African ancestry, aggressive disease biology, and early onset will enable the characterization of the tissue composition and the proportion of cell sub-populations implicated in tumorigenesis and the interplay with disease predisposition and severity.

# Integrated gene expression, and single-cell and spatial transcriptomics, identifies early molecular and cellular events in gastric carcinogenesis

**Andrew Su[1,2] *, Ignacio Wichmann[1, 3] *, Anuja Sathe[1], Jeanne Shen[4], Susan M. Grimes[1] , Xiangqi Bai[1], Quan Nguyen[2], Joo Ha Hwang[5], Robert Huang[5], Hanlee P. Ji[1]**

* These authors equally contributed to this work

1 Division of Oncology, Department of Medicine, Stanford University, School of medicine, Stanford, CA, 94305, United States
2 Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, 4072, Australia
3 Division of Obstetrics and Gynecology, Department of Obstetrics, Escuela de Medicina, Pontificia Universidad Católica de Chile, Santiago, 8331150, Chile
4 Department of Pathology, Stanford University School of Medicine, Stanford, CA, 94305, USA
5 Division of Gastroenterology, Department of Medicine, Stanford University School of Medicine, Stanford, CA, 94305, United States

We developed an integrated multi-omics approach to analyze early cancer precursors. This approach enabled us to infer a richer set of cellular genomic features than would otherwise be generated from a single platform assay. This integrated approach combined conventional RNA-seq of these precursors with information from spatial transcriptomics and single cell RNA-seq. On its own, standard gene expression studies provide molecular signatures which can be associated with specific clinical outcomes such as increased risk of cancer. However, gene expression as a sole data source can neither identify individual aberrant cells, nor provide topographic disease tissue features such as mapping expression signatures to specific precancerous tissue regions. Our approach addressed all of these limitations.

*(continue to page 72)*

*(continued from page 71)*

RNA-seq was applied to an extended cohort of early stomach cancer precursors that are referred to as gastric intestinal metaplasia. We leveraged the sample size of RNA-seq to identify a novel high-risk gene signature. We applied spatial transcriptomics (10X Visium) to an independent set of gastric intestinal metaplasia and gastric cancer samples. This spatial information enabled us to map differentially expressed high-risk genes specifically to pathologist-annotated regions of intestinal metaplasia. To improve the inference of the affected cells types in regions of metaplasia, we used single cell information to make assignment of aberrant epithelial cells. Finally, we used InferCNV on the spatial gene expression data to identify chromosome 8q amplifications, an early event associated with gastric cancers and restricted only to the aberrant epithelial cell types. Overall, our integrated multi-omics approach enabled the identification of early genomic events harbored by aberrant cell types in high-risk precancerous lesions. This approach may prove useful for discovering new clinical and diagnostic biomarkers.

# Molecular characterization of early lung adenocarcinoma based on whole-genome sequencing and spatial transcriptome analysis



**Ayako Suzuki[1], Masayuki Noguchi[2], Takashi Kohno[3], Yutaka Suzuki[1]**

1 Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan
2 Center for Clinical and Translational Science, Shonan Kamakura General Hospital, Kanagawa, Japan.
3 Division of Genome Biology, National Cancer Center Research Institute, Tokyo, Japan.

To characterize molecular events during stepwise progression of lung adenocarcinoma, we performed whole-genome sequencing of 76 non-small cell lung cancer cases, including adenocarcinoma in situ (AIS) and minimally invasive adenocarcinoma (MIA). AIS and MIA are very early stage of adenocarcinoma and correspond to Noguchi type A, B and early C by Noguchi classification.

As a result, driver mutations already occurred even in AIS. Mutations in tumor-suppressor genes were enriched in advanced cases. Interestingly, RBM10 mutations highly occurred in early cases while SMARCA2/SMARCA4 mutations were biasedly detected in advanced cases. We also found that hyper-mutator phenotypes harboring a large number of point mutations and SVs appeared in advanced cases. On the other hand, copy number aberrations highly increased in a part of later early cases, which might be associated with genome-wide hypomethylation statuses called from PromethION data.

To understand patterns of mutation occurrence at a haplotype level, we assigned the detected mutations to each of the two haplotypes that were constructed by haplotype phasing analysis using long read data. We found that mutation-enriched regions in which mutations biasedly occurred in one haplotype showed mutation patterns similar to APOBEC signatures, which were also detected in AIS.

Gene expression analysis revealed that genes associated with immune migration and angiogenesis were highly expressed in AIS Noguchi type B. We also found that more invasive signatures, such as collagen fibril organization, DNA repair and proliferation, were highly expressed in MIA. The results indicated that tumor cells would start interaction with microenvironment in Noguchi type B and they might become to represent malignant signals in early invasive adenocarcinoma. To validate them, we performed spatial transcriptome analysis Visium. We integrated omics and histopathological information to identify key molecular events of lung adenocarcinoma progression at a haplotype, multi-layered and spatial level.

# NGS library prep? We've got you covered.

For 13 years, NEB® has addressed library prep challenges by offering solutions that streamline workflows, minimize inputs, and improve library yield and quality. In fact, use of NEBNext® reagents has been cited in >20,000 publications.

With over 90 products in the NEBNext portfolio and direct access to experts in NGS and enzymology, NEB can support library preparation for all of your sample types, for a wide range of sequencing platforms and throughputs.

The NEBNext Ultra™ II workflow, which lies at the heart of the NEBNext library prep solution, is available in a variety of streamlined and convenient kit formats. For added flexibility, NEBNext kits are easily scalable for automation on liquid handling instrumentation. As a reagent manufacturer, NEB is ideally positioned to support your large volume and custom needs through our OEM & Customized Solutions group.

Our expertise, proven track record and depth of portfolio make NEBNext a preferred choice for high-quality sample preparation. We're ready to show you why it should be yours.

Interested in giving NEBNext a try? Learn more and request a sample at **NEBNext.com**.



NEW ENGLAND BioLabs® Inc.

*be* INSPIRED
*drive* DISCOVERY
*stay* GENUINE

# Information & Computational Biology

# Democratizing access to microbial data and bioinformatics tools



**Françoise Thibaud-Nissen and Nuala O'Leary**

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

The number of prokaryotic genomes submitted to GenBank has increased every year, posing a challenge to these retrieving these data from the NCBI databases and trying to understand what makes a particular genome unique or valuable compared to others.

The RefSeq collection provides consistent and up to date gene annotation for more than 260,000 of these genomes using PGAP, a mature pipeline that incorporates expert-curated reference data. To facilitate the comparison of gene sets across genomes, PGAP is now adding Gene Ontology terms to proteins and protein-coding genes annotated on genomes. A gene annotated on a genome may have a gene symbol, an EC number and/or a GO term in addition to a described function. The GO terms are inherited from the collection of protein family models (PFMs) used by the PGAP functional annotation process and made available in the Entrez Protein Family Models resource (https://www.ncbi.nlm.nih.gov/protfam). In addition to the PGAP annotation itself, an assessment by CheckM (Parks DH et al. 2014. Genome Research, 25: 1043-105) of the completeness of the annotated genome is now performed as part of the process.

For users wishing to annotate their own genomes in the same manner as RefSeq, PGAP (including the CheckM module) is available as a Docker container that can be executed outside of NCBI on an individual computer, a compute farm, or in a cloud environment from an intuitive command-line interface (see https://github.com/ncbi/pgap).

Finally, this presentation will demonstrate how Datasets, a new NCBI resource for accessing sequence data on the command line, via specialized APIs or the web can facilitate the retrieval of prokaryotic assemblies and proteins and of their associated metadata, including GO terms and CheckM completeness scores.

# Genomewide association study in a rat model of temperament identifies multiple loci for exploratory locomotion and anxiety-like traits

**Apurva S. Chitre[1*], Elaine K. Hebda-Bauer[2*], Peter Blandino[2], Hannah Bimschleger[1], Khai-Minh Nguyen[1], Pamela Maras[2], Fei Li[2], A. Bilge Ozel[3], Oksana Polysskaya[1], Riyan Cheng[1], Shelly B. Flagel[2], Stanley J. Watson, Jr.[2], Jun Li[3], Huda Akil[2**], Abraham A Palmer[1,4**]**

*Equal first authors, **Equal last authors, # Corresponding Author

1 University of California San Diego, Department of Psychiatry, San Diego, California, USA
2 University of Michigan, Michigan Neuroscience Institute, Ann Arbor, Michigan, USA
3 University of Michigan, Department of Human Genetics, Ann Arbor, Michigan, USA
4 University of California San Diego, Institute for Genomic Medicine, San Diego, California, USA

Common genetic factors likely contribute to multiple psychiatric diseases including mood and substance use disorders. Certain stable, heritable traits reflecting temperament, termed externalizing or internalizing, play a large role in modulating vulnerability to these disorders. To model these heritable tendencies, we selectively bred rats for high and low exploration in a novel environment (bred High Responders (bHR) vs. Low Responders (bLR)). To identify genes underlying the response to selection, we phenotyped and genotyped 558 rats from an F2 cross between bHR and bLR.

We obtained 4,425,349 Single Nucleotide Polymorphisms (SNPs) across 538 F2 rats using low-coverage whole genome sequencing (lcWGS). Multiplexed sequencing libraries were prepared using the Riptide (Twist Bioscience) library preparation kit and then sequenced on a NovaSeq 6000 platform (Illumina). Because SNPs that segregate between bHR and bLR were not known, we deeply sequenced 20 rats from the F0 generation. We used STITCH to impute genotypes in the F2 using the SNPs from the 20 deeply sequenced F0 rats as reference data. Next, we used BEAGLE to impute missing genotypes. To estimate genotyping accuracy and to identify potential sample mix-ups due to sample handling errors, we compared genotypes to the earlier microarray-derived genotypes (Zhou et al., 2019). We used these dense, high-quality genotypes to perform genetic analysis.

*(continue to page 78)*

*(continued from page 77)*

We identified multiple significant loci for six behavioral traits. Five of the six traits reflect different facets of EL that were captured by three behavioral tests. Distance traveled measures from the open field and the elevated plus maze map onto different loci, thus may represent different aspects of novelty-induced locomotor activity. The sixth behavioral trait, number of fecal boli, is the only anxiety-related trait mapping to a significant locus on chromosome 18. The identification of a locomotor activity-related QTL on chromosome 7 encompassing the Pkhd1l1 and Trhr genes is consistent with our previous finding of these genes being differentially expressed in the hippocampus of bHR vs. bLR rats. In conclusion, our selectively bred rat model reveals greater insight into the genetic architecture of sensation-seeking, anxiety, and addiction-related traits.

# Integrating single-cell and spatial gene expression profiling of mouse organogenesis to identify and localize unknown cell type

**Stephanie Zimmerman[2], Chengxiang Qiu[1], Eva Nichols[1], Alexa Lasley[2], Margaret Hoang[2], Joe Beechem[2], Jay Shendure[1]**

1 Department of Genome Sciences, University of Washington, Seattle, WA, USA
2 NanoString® Technologies, Seattle, WA, USA

Mammalian organogenesis is a remarkable process, whereby cells within the post-gastrulation embryo continue to rapidly proliferate while giving rise to the diverse cell types of each organ system, specified by molecular programs that are precisely regulated in time and space. Single cell RNA-sequencing of whole embryos during mouse embryogenesis and organogenesis is yielding unprecedented detailed views of mammalian development, for example revealing hundreds of unique cell types defined by gene expression. Although many methods exist for identification of cell types defined by scRNA-seq, annotating cells remains a manual and challenging process. In this work, we sought to leverage spatial gene expression data of mouse organogenesis to validate annotations and localize uncertain cell populations to specific tissues or regions.

We used high-resolution scRNA-seq data from mouse development -- specifically, single nucleus transcriptional profiling of millions of cells done by 3-level combinatorial indexing. This “whole organism profiling” was performed on staged embryos in 2 to 6 hour increments from gastrulation to birth. The resulting single cell profiles were processed by conventional methods and, although an initial round of manual annotation based on marker genes and earlier generations of atlases was fruitful, many ambiguities remained. To address these in part, we integrated matched timepoints with spatial whole transcriptome profiles of specific anatomical structures of four timepoints of mid-gestation mouse development generated using the GeoMx® Digital Spatial Profiler (DSP). We used a cell type deconvolution algorithm to estimate the abundance of each cell type in each region profiled by the DSP instrument and validated that known cell types such as tissue-specific epithelial cell subtypes localize to the correct anatomical structures with high accuracy. We then used this method to further map the cell trajectories derived from the lateral plate mesoderm, populations which have limited research and are therefore challenging to annotate. This work provides a framework for integrating spatial data with scRNAseq in an automated pipeline to add spatial context to unknown cell types.

# Repurposing HiSeq 2500 sequencing systems as versatile fluidics and imaging platforms with PySeq2500

**Kunal Pandit[1,5], Joana Petrescu[2,3,4,5], Miguel Cuevas[2,3], William Stephenson[1], Peter Smibert[1], Hemali Phatnani[2,3,4], Silas Maniatis[1]**

1. Technology Innovation Lab, New York Genome Center, New York, NY, USA
2. Department of Neurology, Columbia University Irving Medical Center, New York, NY, USA
3. Department of Applied Physics and Math, Columbia University Irving Medical Center, New York, NY, USA
4. Center for Genomics of Neurodegenerative Disease, New York Genome Center, New York, NY, USA

These authors contributed equally to this work.

The HiSeq 2500 system was a legacy workhorse sequencer, where over 2000 systems were in use as of 2018. Today, HiSeq 2500 systems have been obsolesced by new sequencers and Illumina has announced they will stop supporting the instrument and supplying the reagents in early 2023. However, the sequencer is still capable of performing as a robust, productive, and versatile benchtop system of scientific equipment that includes precise XYZ stages, temperature control, fast imaging, and integrated fluidic handling. We repurposed a HiSeq 2500 system to characterize mouse and human tissue sections with iterative indirect immunofluorescence imaging (4i). The stages, pumps, valves, optics, and cameras of the HiSeq were controlled and automated with our open source software, PySeq2500. Custom flow cells were developed from off the shelf components to mount tissue sections and to interface with the HiSeq fluidics. Only minor modifications to the HiSeq stage were made to accommodate the custom flow cell and to the fluidic plumbing to minimize dead volume. Simple PySeq2500 protocols enable versatile experimental designs in large flow cells involving simultaneous 4-channel image acquisition, temperature control, reagent exchange, and stage positioning. Unattended execution of complex, multi-day workflows was demonstrated by automating 4i with PySeq2500. Our automated 4i method used off-the-shelf antibodies over multiple cycles of staining, imaging, and elution to build highly multiplexed maps of cell types and pathological features across entire mouse and postmortem human spinal cord sections. PySeq2500 enables non-specialists to automate state of the art fluidics coupled imaging methods on widely available HiSeq 2500 systems.

# Measuring cell identity and state transitions during continual biological systems



**Wenjun Kong**, **Calico Life Sciences**

Transitions in cell identity are fundamental to development, reprogramming, disease, and aging. Single-cell genomic technologies enable the dissection of tissue composition on a cell-by-cell basis in complex biological systems. Here, we present a computational tool, 'Capybara,' to classify discrete cell identity and intermediate "hybrid" cell states, supporting a metric to quantify cell fate transition dynamics. We validate hybrid cells using experimental lineage tracing data of hematopoiesis, highlighting the multi-lineage potential of these cell states. Leveraging this approach, we diagnose limitations of various cell fate engineering strategies, revealing previously uncharacterized patterning deficiency, off-target identities, and a putative in vivo correlate for an engineered cell type that has, to date, remained undefined. Further expansion of this approach in joint with multi-modal high-resolution genomics technologies, such as spatial transcriptomics and single-cell multiomics, highlights the deteriorating liver zonation during aging. These applications prioritize interventions to increase efficiency and fidelity of cell engineering strategies and identification of key changes across transitions, showcasing the broad utility of Capybara to dissect cell identity and fate transitions continual biological systems.

# Multiplexing samples for simultaneous measurement of RNA and ATAC in single nuclei: A bioinformatic perspective

**Heather Geiger**, **Rui Fu, Salima Soualhi, Jade Carter**
**New York Genome Center, New York, NY, USA**

Simultaneous measurement of gene expression (RNA-Seq) and chromatin accessibility (ATAC-Seq) at the single cell level and at high throughput may reveal a greater understanding of epigenetic regulation and how it differs by cell type. Here, we aimed to apply a barcoded antibody method (hashing) to a commercially available kit from 10X Genomics, the Chromium Single Cell Multiome ATAC + Gene Expression kit.

The current version of this Multiome kit, which enables users to profile in a single experiment gene expression and chromatin accessibility from isolated nuclei up to 10,000 cells, requires that each sample is processed individually: this is an important limitation which impacts negatively experimental time and cost. Furthermore, this approach contrasts with other protocols (scRNA-Seq, snRNA-Seq, scATAC-Seq) which have been successfully adapted to multiplexing strategies to increase cell loading number capabilities and reduce per-sample cost.

We have validated the application of single-nuclei hashing antibodies to the 10X Multiome protocol using both mouse and human brain samples. We found a high rate of specificity within sample identity calls, as well as similar QC metrics for RNA and ATAC to those generated without hashing. Further, with the aid of multiplex hashing to identify and remove multiplets, cell loading beyond the recommended 10,000 limit can be achieved with little dropoff in data quality.

Consistent with other hashing methods, we do find that for a number of nuclei that sample origin cannot be assigned (“Negatives”), especially for non-neuronal cell types such as oligodendrocytes. Here we demonstrate that a modification of the Seurat HTODemux method, in which all cells are normalized to a fixed number of total HTO counts before running the usual CLR- transformation and HTODemux steps, may reduce the number of negatives while retaining a high specificity rate for singlets. Our recent findings illustrate that this method is efficient for reducing cell type bias in Multiome and single nuclei hashing data. Further, its suitability for other sub-optimal hashing situations are also explored.

# Pacybara: Long-read clustering of barcoded libraries for multiplexed assays of variant effects

**Jochen Weile[1-4], Atina Cote[1-4], Nishka Kishore[1-4], Daniel Tabet[1-4], Ashyad Rayhan[1-4], Frederick P Roth[1-4]**

1 Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto, ON, M5G 1X5, Canada
2 The Donnelly Centre, University of Toronto, Toronto, ON, M5S 3E1, Canada
3 Department of Molecular Genetics, University of Toronto, Toronto, ON, M5S 3E1, Canada
4 Department of Computer Science, University of Toronto, Toronto, ON, M5S 2E4

Multiplexed assays of variant effects (MAVEs) are increasingly used to create comprehensive genotype-phenotype maps in a bid to aid clinical variant interpretation. Many MAVE methods rely on barcoded mutant libraries and thus require accurate lookup tables for the association between barcode and genotype. These tables are often generated via high-fidelity long-read sequencing and specialized processing pipelines. Existing pipelines such as Subassembly and PacRat extract barcode regions for all reads, group them by identical barcodes, and form them into consensus sequences.

However, a number of limitations exist with this approach: (1) A single barcode sequence may have been associated with multiple independent clones within a given library; (2) Sequencing errors in the barcode region can cause a single clone to appear as two or more distinct clones, or vice versa; and (3) PCR crossover events that may have been introduced during library preparation can yield consensus genotypes that are chimeric with wild-type or other variant genotypes.

We have developed Pacybara—a long-read clustering tool, which addresses the aforementioned limitations. Pacybara supports amplicons with multiple barcode sites and takes into account sequencing errors within the barcodes. Using an iterative process, clusters at increasing distance in sequence space are considered for merging. To adjudicate a proposed merge, Pacybara considers the total number of divergent base calls between clusters, their respective quality scores, their Jaccard coefficient, and the relative cluster size difference.

We demonstrate that Pacybara successfully detects non-unique barcode and PCR crossover events, reduces false positive indel calls, and increases the sensitivity of the resulting variant effect map.

# Prenatal screening for pathogenic submicroscopic copy number variations in circulating fetal cells from maternal blood

**Martina Dori[1], Anna Doffini[1], Chiara Mangano[1], Claudio Forcato[1], Roberta Aversa[1], Chiara Maranta[1], Emilia D. Giovannone[1], Genny Buson[1], Chiara Bolognesi[1], Rebecca Maiocchi[1], Debora Lattuada[2], Enrico Ferrazzi[2], Paola Ricciardi-Castagnoli[1] Thomas J. Musci[1] and Francesca Romana Grati[1].**

1 A. Menarini Biomarkers Singapore Pte Ltd, Singapore
2 Department of Woman Child and Neonate, Obstetrics Unit, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy

The analysis of cfDNA demonstrated the potential for non-invasive prenatal testing (NIPT) for common aneuploidies screening, but is limited for syndromes involving micro-imbalances, which requires a genome-wide high-resolution analysis. The isolation of intact circulating ExtraVillous Trophoblasts (cEVT) from maternal blood represents a promising approach for non-invasive prenatal testing, as the analysis of single cells can overcome the constraints of other cfDNA-based approaches.

Here we present an automated approach for the cEVTs isolation and the performance of our NGS-based microdeletion analysis on single cells of cell-lines (Coriell CNV Reference) harboring well-characterized pathogenic copy number variants (pCNVs) and on cEVTs isolated from maternal peripheral blood enrolled during a proof-of-concept (POC) study.

Enrichment and isolation of cEVTs from maternal blood was done using a novel automated antibody-based capture method and isolation of single cells on an image-based sorter (DEPArrayTM). Fetal single cells were confirmed by maternal-to-fetal microsatellite comparison and analyzed by NGS on whole-genome amplified DNA for CNVs. Coriell genomic DNAs reference panel were used to assess the single-cell NGS workflow limit of detection (LoD) for micro-imbalance. Moreover, single cells from 5 cell lines of this panel were isolated by flow cytometry and analyzed using the same bioinformatic pipeline used for cEVTs.

*(continue to page 85)*

*(continued from page 84)*

Experiments with 43 individual gDNA samples with known pCNVs showed that our current limit of CNV size detection is ~0.8-1Mb. Through the isolation of 73 single cells from 5 different cell-lines of the Coriell CNV Panel, we demonstrated a consistent reproducibility of the method (Positive Percentage Agreement ⩾90% in both single cells and gDNA) with microdeletions/duplications of 0.8Mb in size. Notably, all isolated cells from GM17942 cell line showed high-quality genomic profiles, which allowed us to consistently identify the expected 22q11.2 deletion (DiGeorge Syndrome). In cases enrolled through the POC study we were able to detect a pathogenic deletion on chr16 of 0.8Mb in size in cEVTs, also validated by microarray analysis after invasive procedure.

Our overall findings suggest that the isolation of cEVTs coupled with NGS-based approach may allow a novel opportunity for a reliable comprehensive non-invasive high-resolution fetal genome profiling in prenatal testing, particularly for detection of pathogenic microdeletion/duplication.

# Shining light on the dark genome: Accurate quantification of tandem repeats for translational and basic research

**Egor Dolzhenko[1], Adam English[2], William J. Rowell[1], Zev Kronenberg[1], Dalia Kasperaviciute[3], Henry Houlden[4,5], Roisin Sullivan[4,5], Jana Vandrovcova[4,5], Arianna Tucci[3], Genomics England Research Consortium[3], Aaron M. Wenger[1], Fritz J. Sedlazeck[2,6], Michael A Eberle[1]**

1 Pacific Biosciences of California, Menlo Park, California, USA
2 Baylor College of Medicine, Houston, Texas, USA
3 Genomics England Limited, Dawson Hall, Charterhouse Square, London, UK
4 National Hospital for Neurology and Neurosurgery, Neurogenetics Unit, London, UK
5 University College London, Institute of Neurology, Department of Neuromuscular Diseases, London, UK
6 Rice University, Department of Computer Science, Houston, Texas, USA

Tandem repeats (TRs) are genomic regions consisting of exact or near-exact repetitions of DNA sequence. In addition to being associated with many genetic disorders and gene expression levels, up to 80% of structural variants occur within TR regions, underscoring the need to study TR variation across the genome. TRs are difficult to analyze with short-read sequencing data due to their repetitiveness and structural complexity. Because of this, standard computational methods for variant calling in TR regions often fall short of capturing the true allele sequences. Here we describe the first method for accurate genome-wide characterization of length, motif variation, and methylation of tandem repeats called the Tandem Repeat Genotyping Tool (TRGT). Our benchmarks revealed that genotypes produced by TRGT have high Mendelian concordance (over 99% percent of analyzed repeats were concordant in family trios) and that TRGT can accurately detect known repeat expansions in FMR1, HTT, ATXN8, RFC1 and other genes. We used TRGT to analyze the genetic and epigenetic variation of almost one million TR regions in 100 PacBio HiFi whole-genome samples. We summarized the distributions of lengths, sequence composition, and DNA methylation levels into a database and developed a set of tools for querying and adding new data to this database.

*(continue to page 87)*

*(continued from page 86)*

We also enabled TRGT to genotype a novel class of highly variable repeats that cannot be described by a fixed-length consensus sequence. These TRs correspond to clusters of variation (often repeats within repeats) that can produce spurious variant calls. We show that the TRGT probabilistic framework allows us to accurately represent the structure of these repeats, making it possible to analyze them as any other TR in the human genome. Taken together, our results shed light on the structural complexity and population variation of TRs across the genome. Accurate TR analysis improves our ability to detect known pathogenic expansions and discover novel functional repeats.

# Solving the puzzle of pediatric genetic disease with bits and bytes



**Ben Kelly and Peter White**

Genomic medicine positively impacts pediatric care, from rapid diagnosis of genetic disorders to optimization of childhood cancer treatments. The Computational Genomics Group at Nationwide Children's Hospital combines genomics and bioinformatics to improve patient outcomes by supporting multiple translational research protocols and developing research ideas into clinical applications. Our Genomics of Rare Disease protocol has utilized genome sequencing and novel bioinformatics approaches to find answers for patients with undiagnosed disease, revealing novel disease mechanisms. Through the application of cutting-edge cloud computing technologies, optimized bioinformatics pipelines, and an Apache Spark-backed variant and annotation warehouse, our Rapid Genome Sequencing protocol returns phenotype-driven results for infants in the ICU within 48 hours. Our Cancer protocol provides extensive genomic and transcriptomic profiling, positively impacting diagnosis, prognosis, and therapy for patients with rare and recurrent disease. Finally, the Molecular Characterization Initiative is a new screening approach to identify therapeutic vulnerabilities in pediatric cancers, by performing genomic, transcriptomic and epigenomic characterization of thousands of pediatric cancer samples nationwide. All genomic data is generated, analyzed, and interpreted within 14 days and deidentified genomic and clinical data is submitted to dbGaP within minutes of a case being completed. Together with sharing data from our translational protocols, we are creating a community resource that will enable wide-scale engagement of the translational bioinformatics community to help solve the puzzle of genetic disease in pediatrics and beyond.

# Understanding genomic variation using Glassbox machine learning



**Nathan Dwarshuis**
NIST

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 specific, quantifiable features of the context surrounding a given variant.

We aim to make such a model using explainable boosting machines (EBMs). EBMs create models that are a linear combination of arbitrary univariate and bivariate functions (a generalized additive model with interaction terms). Despite being flexible, the relative simplicity of EBMs allows a human to easily understand the functional relationship and relative contribution of each feature. We exploit this transparency here by fitting EBMs to variant call errors as a function of genomic context, which enables understanding when and to what extent a given feature will impact performance.

We demonstrated this strategy with two main use cases. First, we compared false positive rates for Illumina PCR-free and PCR-plus sequencing technologies. Within homopolymers, INDEL false positives were much more prevalent in PCR-plus, and that the error rate for both methods began to increase after 8 bp for both A/T and G/C homopolymers. SNP errors in contrast were not different between the two technologies; however, the error rate increased beyond 10 bp for A/T homopolymers and any length in the case of G/C homopolymers. For our second use case, we compared the likelihood of callsets from Illumina PCR-free or PacBio HiFi data to miss clinically relevant variants. Most false negatives occurred in hard-to-map regions, but some errors occurred in homopolymer regions.

Ultimately, this will provide a data-driven mechanism for understanding sources of error within different variant caller methods and sequencing technologies as they relate to calling variants in difficult regions of the genome. We also plan to use this model to improve genome stratifications for creating and using GIAB benchmarks.

# Visualization and analysis of nanopore RNA and DNA signal alignments for modification detection and more with Uncalled4

**Sam Kovaka[1], Vikram Shivakumar[2], Katharine Jenike[3], Paul W. Hook[2], Luke Morina[2], Roham Razaghi[2], Winston Timp[2,3], Michael C. Schatz[1,3]**

1 Department of Computer Science, Johns Hopkins University, Baltimore, MD
2 Department of Biomedical Engineering, Johns Hopkins School of Medicine, Baltimore, MD
3 Department of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD

RNA has a wide range of natural nucleotide modifications, with over 150 modifications comprising the epitranscriptome. Most of these are poorly understood, in large part because they are difficult to detect with conventional methods. While nanopore sequencing has the potential to identify any nucleotide modification from native DNA and RNA, modern basecallers do not have the ability to directly detect any RNA modifications, and only a handful can be directly basecalled in DNA (5mC, 5hmC). A common way to detect modifications in nanopore data is to align the raw signal to a nucleotide reference. This method, also known as "resquiggling" in Tombo or "eventalign" in Nanopolish, is accomplished by translating a nucleotide reference into an expected signal using a k-mer pore model and performing a signal-level alignment using a dynamic programming algorithm. Signal alignment has been used for modification detection, adaptive sampling, assembly polishing, and more. Despite its widespread use, little attention has been given to comparison and optimization of signal alignment algorithms, and these methods have not been applied to the latest R10.4 pore chemistries, largely due to the lack of a k-mer model.

Here we present Uncalled4: a toolkit for nanopore signal alignment, analysis, and visualization. Uncalled4 alignments outperform Nanopolish and Tombo in detection of human m6A RNA modifications and in Escherichia coli rRNA, both using builtin comparison statistics and when input to third-party tools like xPore and Nanocompore. Multiple input and output formats are supported Signal alignments can also be used to iteratively train new pore models, which allows for alignment of reads generated with the latest R10.4 DNA pore chemistry. The k-mer model alone reveals interesting characteristics of the R10 pore, for example that the nucleotide at the 3' reader-head dominates the signal. Uncalled4 also features a variety of visualizations of signal alignments which can be viewed and manipulated interactively using a browser-based application. Uncalled4 is implemented in C++ and Python and is available at skovaka.github.io/UNCALLED.

# Application of an in silico framework to map genetic networks and elucidate biological functions of KMT2D, a frequently mutated gene across cancer types

**Yuka Takemon[1,2], Alessia Gagliardi[2], Susanna Y. Chan[2], Diane L. Trinh[2], James T. Topham[2], Ryan, D. Huff[3], Christopher S. Hughes[4], Marco A. Marra[2,5]**

1 Genome Science and Technology Graduate Program, University of British Columbia, Vancouver, BC, Canada
2 Canada's Michael Smith Genome Sciences Centre, BC Cancer Research Institute, Vancouver, BC, Canada
3 Air Pollution Exposure Laboratory, Division of Respiratory Medicine, Department Medicine, Vancouver Coastal Health Research Institute, University of British Columbia, Vancouver, BC, Canada
4 Department of Molecular Oncology, BC Cancer Research Institute, Vancouver, BC, Canada
5 Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada

Genetic interaction (GI) networks and essentiality networks have been leveraged to identify therapeutic opportunities in cancers and decipher genes' functional memberships. The Cancer Dependency Map (DepMap) public data platform is the largest resource to mine GI and essentiality networks derived from human cancer cell lines. Although several novel algorithms utilising DepMap data to predict GIs have emerged (eg. De Kegel et al., 2021; Benfatto et al., 2021), these tools either lack flexibility, do not provide the ability to detect fitness-promoting GIs, or require advanced programming knowledge. Thus, more accessible tools would benefit the broader research community. To address this gap, we created a computational tool called Genetic inteRaction and EssentialiTy mApper (GRETA) that provides a simplified approach to query and select DepMap cancer cell lines, based on user-defined genetic features, and identifies DepMap gene knockouts that may be lethal or advantageous in the context of those particular cell lines. We also adapted a previously published method (eg. Kim et al., 2019) to map essentiality networks of a user's gene of interest. Here, we applied GRETA to predict GI and essentiality networks of KMT2D, which encodes a member of the COMPASS complex, and to demonstrate GRETA's utility in deciphering functional associations.

*(continue to page 92)*

*(continued from page 91)*

Somatic KMT2D alterations have been found in several cancer types, including follicular lymphomas (89%), diffuse large B-cell lymphomas (32%), urothelial bladder carcinomas (27%), and gastric cancers (12%). Loss of KMT2D has been associated with an increase in transcriptional stress and genomic instability; however, we do not yet understand how KMT2D influences these processes. Using GRETA, KMT2D-deficient cell lines, and DepMap data, we constructed an essentiality network, which revealed expected associations with other known COMPASS subunits, and with genes that have roles in DNA recombination. We also discovered an apparent synthetic lethal interaction between KMT2D and WRN, which encodes a protein with both helicase and nuclease activities and plays a role in homology-directed repair and replication fork maintenance. Together, our results indicate that KMT2D may play a role in DNA replication and repair, and highlight how GRETA can be applied to understand functional associations of genes that are altered in cancers.

**Abstract Keywords:**
Genetic networks, genetic interaction, gene essentiality, in silico genetic screens

# Multiscale analysis of pangenome enables improved representation of genomic diversity for repetitive and clinically relevant genes

**Chen-Shan Chin, Sairam Behera, Asif Khalak, Fritz J Sedlazeck, Justin Wagner, Justin M. Zook**

Pangenomes are assemblies of a population of genomes, which have the potential to uncover biologically important forms of genomic variation. Improvements to genome assembly contiguity, including the first telomere-to-telomere human genome, have unlocked the potential for applications to precision medicine. We develop a pangenome research toolkit (PGR-TK, https://www.biorxiv.org/content/10.1101/2022.08.05.502980v2 ) enabling analyses of complex genomic variations over a pangenome reference. Such variations include a range of scales from small tandem repeats to ampliconic gene arrays and even megabase scale re-arrangements. A key algorithmic advancement is using a multi-level minimizer index for identifying homologous sequences in the pangenome to construct the minimizer anchored pangenome graph. With such a sparse and efficient index data structure, we can construct pangenome graphs with minimum computation resources from 47 fully assembled and haplotype resolved human genomes released by the Human Pangenome Reference Consortium. It allows researchers to study genomic regions of interest with the flexibility of constructing pangenome graphs with different scales of detail. We introduce a graph decomposition method that automatically identifies different elements into principal bundles of the pangenome. It allows researchers to visualize and interpret haplotypes of complicated genomic structures, thereby avoiding onerous manual processing. Surveying a set of 395 clinical and medically important genes provides quantitative insights into repetitiveness and diversity that could impact the accuracy of variant calls. We apply the graph decomposition methods on the Y-chromosome gene, DAZ1/2/3/4 whose structural variants have been linked to male infertility, as well as X-chromosome genes OPN1LW and OPN1MW. Moreover, we utilize the principal bundle decomposition to understand the highly polymorphic major histocompatibility complex class II region to understand the relationship between the structural variants and the diversity of the HLA-DRA an HLA-DRB genes in the selected pangenome populations.

# Establishing a *Solanum* pan-genome to dissect dynamics of paralog evolution

**Katharine Jenike[1*], Matthias Benoit[2,3*], Srividya Ramakrishnan[1], Shujun Ou[1], James Saterlee[3], Iacopo Gentile[3], Anat Hendelman[3], Michael Passalacqua[3], Xingang Wang[3], Michael Alonge[1], Hamsini Suresh[3], Ryan Santos[3], Blaine Fitzgerald[3], Gina Robitaille[3], Edeline Gagnon[4], Melissa Kramer[3], Sara Goodwin[3], W. Richard McCombie[3], Jaime Prohens[5], Tiina E. Särkinen[6], Amy Frary[7], Jesse Gillis[3,8], Joyce Van Eck[9,10], Zachary B. Lippman[2,3†], Michael C. Schatz[1†]**

1 Computer Science, JHU, Baltimore, MD, USA
2 Howard Hughes Medical Institute, Cold Spring Harbor, NY, USA
3 CSHL, Cold Spring Harbor, NY, USA
4 Technical University of Munich, Chair of Phytopathology, Germany
5 Universitat Politècnica de València, València, Spain
6 Royal Botanic Garden Edinburgh, Edinburgh, UK
7 Mount Holyoke College, South Hadley, MA, USA
8 University of Toronto, Toronto, Canada
9 Boyce Thompson Institute, Cornell University, Ithaca, NY, USA
10 Plant Breeding and & Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, USA

* Authors contributed equally
† Corresponding authors

Pan-genomics and genome engineering of major crops are revolutionizing modern agriculture. A new frontier to strengthen the diversity and resilience of our food system is to improve minor crops. However, insufficient reference genomes and as yet unknown species-specific variation are impeding the translation of knowledge to orphan crops. Gene duplication and paralog diversification create a reservoir of natural variation, but these events are difficult to detect without reference genomes. Moreover, highly heterozygous assemblies are prone to false duplications, complicating paralog analysis in non-model systems.

*(continue to page 95)*

*(continued from page 94)*

Here, we test the hypothesis that paralog diversification is highly dynamic over short evolutionary time scales (~16 MYA), with a goal of revealing how variation in paralogs shapes species-specific phenotypes. We present a pan-genome composed of over 20 ecologically and agriculturally diverse species, including both indigenous crops, sometimes known as minor crops, and wild relatives, spanning the crop-rich Solanum genus. By combining high-fidelity PacBio sequencing with Bionano optical mapping and Hi-C we have produced reference-quality assemblies for each species.

Our *Solanum* pan-genome enables exploration of gene conservation and core sequences within the context of structural variation and transposable element expansion. We collated alleles across the genus within our pan-genome, with a focus on variation within and near paralogous genes, to identify patterns of evolution and address their functional relevance by genome editing. To explore the pan-genome k-mer space, we developed an interactive pan-genome viewer Panagram (https://github.com/kjenike/Panagram). We used Panagram to identify highly similar and highly divergent regions and to quickly view relationships between these distantly related species from the whole genome level to an individual nucleotide resolution. This revealed the landscape of paralog diversification, including a complex history of variation affecting a regulator of fruit size that we dissected using gene editing in the lineage of the minor crop Ethiopian eggplant. Overall, our genus-wide pan-genome provides a new understanding of how paralog diversification impacts trait evolvability and is a foundation to improve predictable translation of genotype-to-phenotype relationships between crops.

# Morphogenomics: Combining nuclear morphology with gene expression

**Jagadish Sankaran[1], Giovani Claresta Wijaya[1], Joseph Lee Jing Xian[1], Vaidehi Krishnan[2], Nirmala Arul Rayan[1], Kok Hao Chen[1], Hein Than[2,3], Sin Tiong Ong[2], Shyam Prabhakar[1]**

1 Genome Institute of Singapore, Singapore, Singapore 138672
2 Duke-NUS Medical School, Singapore, Singapore 169857
3 Department of Haematology, Singapore General Hospital, Singapore 169856

Nuclear dysmorphology coupled with aberrations in gene expression are hallmarks of blood cancers. Although morphological parameters are routinely used in the classification of leukemias, such assessments are based on visual examination and interpretation by the hematopathologist. Multiplexed fluorescence in situ hybridization (mFISH) is a revolutionary technology providing information not only on the expression of hundreds of genes but also on cell and nuclear morphology. Here, we propose MOGEC (morphology gene expression correlation) analysis to investigate the relationships between nuclear dysmorphology and gene expression in blood malignancies using liquid and bone marrow biopsies. Conventionally, in single cell RNA sequencing, clustering is performed only on gene expression to identify cell types. In contrast, MOGEC characterizes cells in a hybrid space that combines both gene expression and shape metrics inferred from mFISH data.

Our implementation of MOGEC was based on an mFISH assay that currently profiles 492 genes in ~25,000 cells in a 16 mm2 region of interest in 34 hours. We applied this approach to peripheral blood mononuclear cells (PBMCs) and parametrized cell and nuclear morphology using 4 metrics: aspect ratio, circularity, roundness, and solidity. The quantified gene expression levels correlated well with estimates from bulk RNA-seq (log-log correlation coefficient=0.6) and enabled identification of major cell types in peripheral blood mononuclear cells (PBMCs).

*(continue to page 97)*

*(continued from page 96)*

Heterogeneity in nuclear shape is a characteristic feature of PBMCs. T and B cell nuclei are circular, whereas the nuclei of monocytes have diverse shapes. Morphology-informed clustering identifies two distinct monocyte populations that could not be distinguished on the basis of gene expression alone. One population has circular nuclei, whereas the other is enriched for non-circular nuclei (primarily heart-shaped). These two MOGEC clusters show subtly distinct gene expression patterns, indicating that they may represent distinct transcriptomic states.

Morphological heterogeneity of monocytes is a prognostic feature of chronic myelomonocytic leukemia (CMML). However, morphological analysis alone may lead to misclassification of immature or dysplastic granulocytes as monocytes. Combining morphological analysis with gene expression using MOGEC could improve CMML patient stratification and reveal the underling biology of clinically relevant malignant cell morphologies.

# Your Science Can't Wait

Element Biosciences

*COMING SOON*

## Avidity Cloudbreak™

- Faster run times
- Simplified workflow with linear library loading
- Industry leading quality
- Lowest run costs at $5/Gb


Learn more at **Lounge #212**

# Medical Omics

# Bronchial gene expression alterations associated with radiographic bronchiectasis

**Yuriy O. Alekseyev[1], Ke Xu[1], Alejandro A Diaz[2], Fenghai Duan[3], Minyi Lee[1], Xiaohui Xiao[1], Hanqiao Liu[1], Gang Liu[1], Michael H. Cho[2], Adam C. Gower[1], Avrum Spira[1], Denise R Aberle[4], George R Washko[2], Ehab Billatos[1], Marc E. Lenburg[1]**

1 Boston University School of Medicine, Boston, MA.
2 Brigham and Women's Hospital, Boston, MA.
3 Brown University School of Public Health, Providence, RI.
4 David Geffen School of Medicine at UCLA, Los Angeles, CA

Discovering airway gene expression alterations associated with radiographic bronchiectasis (BE) may improve the understanding of the pathobiology of early-stage BE. Presence of radiographic BE in 173 individuals without a clinical diagnosis of BE was evaluated. Bronchial brushings from these individuals were transcriptomically profiled and analyzed. Single-cell deconvolution was performed to estimate changes in cellular landscape that may be associated with early disease progression. 20 participants have widespread radiographic BE (>= 3 lobes). Transcriptomic analysis reflects biologic processes associated with BE including decreased expression of genes involved in cell adhesion and increased expression of genes involved in inflammatory pathways (655 genes FDR < 0.1, log2 fold-change > 0.25). Deconvolution analysis suggests that radiographic BE is associated with an increased proportion of ciliated and deuterosomal cells, and a decreased proportion of basal cells. Gene expression pattens separated participants into three clusters: normal, intermediate, and bronchiectatic. The bronchiectatic cluster was enriched by participants with more lobes of radiographic BE ($p < 0.0001$), more symptoms ($p = 0.002$), higher SERPINA1 mutation rates ($p = 0.03$), and higher CT-derived BE scores ($p < 0.0001$). Genes involved in cell adhesion, WNT signaling, ciliogenesis, and IFN-γ pathways had altered expression in the bronchus of participants with widespread radiographic BE, possibly associated with decreased basal and increased ciliated cells. This gene expression pattern not only is highly enriched among individuals with radiographic BE, but also associated with airway-related symptoms in those without discernable radiographic BE, suggesting that it reflects a BE-associated but non-BE specific lung pathophysiogic process.

# Cas9 targeted long-read nanopore sequencing of ABCA4 and the OPN1LW/OPN1MW gene array for investigation of inherited retinal disease

**Gavin Arno[1,2,3] , N. Jurkute[1,2], F. L. Motta[1,4], B. McClinton[5], C. Boles[6], O. A. Mahroo[1,2], M. Michaelides[1,2], A. R. Webster[1,2]**

1 UCL Inst. of Ophthalmology, London, United Kingdom
2 Moorfields Eye Hosp., London, United Kingdom
3 North Thames Genomic Lab. Hub, Great Ormond Street Hosp. for Children, London, United Kingdom
4 Dept. of Ophthalmology, Univ.e Federal de Sao Paulo, Sao Paulo, Brazil
5 Leeds Institute of Medical Research, University of Leeds, St James's University Hospital, Leeds, United Kingdom
6 Sage Sci. Inc., Beverly, MA

Purpose: Short, paired-end sequencing (next generation sequencing, NGS) enables interrogation of up to ~95% of the human genome. Several clinically relevant genes for inherited retinal disease (e.g. OPN1LW/OPN1MW) and variants (e.g. complex re-arrangements) cannot be resolved by NGS technology including whole genome sequencing (WGS). We sought to investigate the use of Cas9 targeting of chromosomal segments (CATCH) using the SageHLS system (SAGE Science) with Oxford Nanopore Technologies (ONT) long-read sequencing for a WGS-intractable suspected ABCA4-retinopathy case and to sequence the NGS-intractable OPN1LW/OPN1MW gene array.

Methods: CATCH was performed for ABCA4 and the OPN1LW/OPN1MW array using CRISPR guide RNAs flanking the regions of interest followed by separation of high molecular weight DNA fragments using the SageHLS system. ONT MinION long-read sequencing was performed on resultant enriched DNA fragments.

Sequencing of the OPN1LW/OPN1MW gene array following CATCH enrichment for a control female subject generated a maximal read depth of 36x, and for a control male, 25x read depth, including full-length reads of up to ~233kb spanning the entire target from guide to guide enabling accurate haplotyping of the opsin arrays.

*(continue to page 102)*

*(continued from page 101)*

Conclusions: This study demonstrates that CATCH-nanopore sequencing is effective for cases intractable to NGS. We were able to generate significant enrichment for target regions and ultra-long reads spanning the entire targets. This study identified a complex structural rearrangement, missed by short-read WGS, in ABCA4 that may lead to cryptic splicing. Effective sequencing of the NGS-intractable OPN1LW/OPN1MW gene array is possible for variant detection in males with blue cone monochromacy.

# Complete resolution of genes with highly homologous gene family members or pseudogenes using long-read PacBio HiFi sequencing



**Xiao Chen[1], Michael Eberle[1]**

1. PacBio, Menlo Park, CA, USA

While whole-genome sequencing (WGS) allows identification of clinically relevant variants in ~90% of the genome, there exist difficult regions and variant classes that remain challenging for short read sequencing. Many medically relevant genes fall into these so-called dark regions where accurate analysis is hindered by the presence of highly similar paralogs, including pseudogenes. Furthermore, high sequence homology promotes unequal crossing over, resulting in frequent copy number variants. This challenges the diploid variant calling model when there are more than two copies of a gene, and silent carriers with two copies of a gene on one chromosome and zero on the other are often miscalled as a non-carrier. PacBio HiFi long-read sequencing is ideal for resolving regions with high sequence homology, but informatics methods are still lacking for segmental duplications that are longer than the HiFi read length.

We developed a HiFi WGS-based informatics method that accurately assembles all haplotypes of highly homologous genes and pseudogenes of the same family, and performs variant calling on each haplotype against the gene of interest. Within this framework, both genes and paralogs/pseudogenes can be annotated according to functional status rather than physical location, enabling identification of the copy number of the functional gene and determination of disease or carrier status. We applied this method to resolve 10 clinically relevant genes with paralogs/pseudogenes including SMN1 (spinal muscular atrophy), CYP21A2 (21-hydroxylase-deficient congenital adrenal hyperplasia), and PMS2 (Lynch syndrome). We assessed these genes in 432 samples across five ethnic populations and among all findings, we: 1) identified major SMN1 and SMN2 haplotypes, and a common two-copy SMN1 allele that could greatly boost the accuracy of silent carrier testing, 2) identified a CYP21A2 allele harboring a stop-gain variant in cis with a duplicated copy of CYP21A2, which is often missed by other technologies and 3) characterized the highly frequent gene conversions between PMS2 and its pseudogene.

*(continue to page 104)*

*(continued from page 103)*

Our method, combined with highly accurate long reads, provides a single framework for resolving a class of the most difficult, clinically important genes, thus bringing us one step closer to consolidating the numerous currently offered genetic tests into a single test.

# CRISPR-Cas9 depletion of high-expression genes in human fibroblast samples increases the diagnostic potential for rare disease

**Jon Armstrong[2], R. I. Corona[1], L-K. Wang[1], A. Y. Huang[1], D. Ramachandran[2], J. Diaz[2], M. F. Khalid[2], O. Abinader[2], A. Siddique[2], K. Brown[2], S. F. Nelson[1]**

1 David Geffen Sch. of Med., Univ. of California-Los Angeles, Los Angeles, CA
2 Jumpcode Genomics, San Diego, CA

Whole genome sequencing (WGS) provides a significant improvement in identification of rare variants over whole exome sequencing (WES). However, this increase in rare variant identification provides only modest enhancement in the diagnostic rate for rare genetic disorders. Whole transcriptome RNA sequencing (RNA-seq) has been used in conjunction with WGS to determine functional consequence of mutations and improve clinical interpretation of rare variants identified in WGS data. This combined approach provides a 15-30% improvement of diagnostic yield for rare genetic disorders, primarily through the observation of rare DNA variants that cause aberrant mRNA splicing. However, the utility of RNA-seq is limited because the data primarily consists of signals from highly expressed genes. Although 80% of disease-associated genes (OMIM) are seen in RNA-seq data from fibroblasts and blood, only 20% of these genes are expressed robustly enough for examination. The remaining OMIM genes are expressed at levels too low for comprehensive analysis. To increase resolution of low-expressed OMIM genes and provide the potential of increasing rare disease diagnosis, we have developed a depletion technology (CRISPRclean) that leverages the CRISPR-Cas9 system to remove library fragments from highly expressed genes prior to sequencing and increase representation of lower expressed genes. The CRISPRclean method utilized 353,628 guides targeting 4450 genes (TPM > 30) observed in RNA-seq data from untreated fibroblast libraries. Our results show a 5-fold increase in representation of low-expressed genes in CRISPRclean-depleted human fibroblast samples.

*(continue to page 106)*

*(continued from page 105)*

In addition, we demonstrate an expansion of 4477 genes at TPM > 30 in depleted samples, increasing the number of high confidence genes for assessing rare splicing aberrations or allelic expression alterations while expanding the interpretability of WGS data. In addition, we show high gene expression correlation across technical replicates in depleted samples, within both targeted and non-targeted genes, providing the power to observe genetic aberrations. The CRISPRclean method can be applied to many tissue types and is easily customized. The method can be applied to pre-constructed sequencing libraries or during the library construction process to increase the dynamic range of useful gene expression and alternative splicing data from any tissue.

# Inspiration4: Understanding changes in DNA composition during spaceflight—a comparative analysis using Inspiration4 data and the NASA Twins Study



**Sebastian Garcia Medina,** **Eliah G. Overbey, Namita Damle, Deena Najjar, Krista Ryon, JangKeun Kim, Matt MacKay, Lior Braunstein, Jaden J. Hastings, Evan E. Afshin, Sean Mullane, Amran Asadi, Jaime Mateus, Susan Bailey, Kelly Bolton, Christopher E. Mason**

Radiation is one of the most significant risks for human spaceflight, in particular during long-duration missions that will traverse beyond the protective Van Allen belts. As such, it is imperative to define the effects of radiation and microgravity on the human genome for the purpose of long exposure to the space environment, such as missions to Mars and permanent Lunar outposts. The NASA Twin Study was the first to define the integrated physiological consequences of long duration spaceflight; ranging from telomere length, genomics, and proteomics to immune function and the microbiome. The study shed light on the chronic inflammatory state resulting from long duration spaceflight and provided the first measures of Clonal Hematopoiesis (CH) in astronauts. CH (also called CHIP) is a condition in which certain Hematopoietic stem cells (HSCs) gain a clonal advantage through acquisition of somatic mutations, and evidence of CH was found in both career–astronaut twins. Mutations in epigenetic regulators, such as DNMT3A and TET2, were found at increased rates when compared to their age-controlled counterparts. To further examine this trend, we collected samples from the Inspiration4 mission (I4), a three-day, all-citizen orbital flight. Since none of the I4 crew had ever been exposed to space, we could define the degree to which short duration spaceflight changes the DNA composition of astronauts and the reversibility of said changes. I4 astronauts presented a transient increase to their telomeres' length as a function of spaceflight, solidifying previous findings about cellular oxidative stress responses in space. Conversely, CH-associated mutations were not disproportionately found in the I4 astronauts post-flight. However, longitudinal tracking of their genome, for years after their flight, could help us determine the long-term effects of spaceflight on development of aberrant mutations and cellular changes in DNA risk alleles.

# Mitochondrial genomics using nanopore Cas9-targeted sequencing



**Amy R. Vandiver[1], Brittany Pielstick[2], Austin Hoang[3], Jonathan Wanagat[3], Winston Timp[2]**

1 David Geffen School of Medicine at University of California, Los Angeles, Department of Medicine, Division of Dermatology, Los Angeles, California, USA
2 Johns Hopkins University, Department of Biomedical Engineering, Baltimore, Maryland, USA
3 David Geffen School of Medicine at University of California, Los Angeles, Department of Medicine, Division of Geriatrics, Los Angeles, California, USA

The mitochondrial genome (mtDNA) is an extranuclear source of genetic variability that regulates critical cell functions, however it is understudied due to the limitations of next generation sequencing (NGS). Mitochondria are considered the "powerplants" of eukaryotic cells, generating up to ninety percent of cellular ATP. Each mitochondrion contains hundreds to thousands of copies of a 16,569 base pair circular genome which is essential for its function. MtDNA is highly susceptible to mutagenesis, including large deletions. When a mutation is present, it can affect all copies, or only a fraction of the mtDNA. This fraction is referred to as the level of heteroplasmy. High mutation heteroplasmy has been linked to tissue dysfunction, cancer and functional decline. While this genome offers a potential metric of disease and target for intervention, mapping and quantifying mtDNA variants has been limited by the challenges of enriching for mtDNA and analyzing a heteroplasmic genome with NGS.

We developed a method for long-read mtDNA sequencing utilizing Cas-9 based adaptor ligation with nanopore sequencing (nCATS). Using this method, we achieve up to 97% enrichment of mtDNA, with up to 66% of reads greater than 15 kbp. These data provide the ability to phase heteroplasmic mutations, analyze population dynamics of mtDNA and map rare structural variants. To assess the capacity for phasing of rare mutations, we applied this method to a blood sample from a patient with an mtDNA-linked syndrome, demonstrating the capacity to phase genomically distant mutations and identify three distinct mtDNA-mutant populations. Next, to develop the capacity for structural variant calling, we used long read mtDNA sequencing to map age-associated deletions across 15 muscle and 6 substantia nigra samples spanning ages 20-81. We observed a strong correlation between mtDNA deletion frequency and age (R2=0.83) and an expanded scope of deletion mutations compared to those previously reported. NCATS-based sequencing of mtDNA thus offers a novel opportunity to understand this important source of genetic variation.

# Transcriptional signature of Neuro–COVID-19 in distinct brain regions

**Nicholas Pasternack*[1], Myoung-Hwa Lee+[1], Wei Yang[2], Daniel P. Perl[3], Joseph Steiner[1], Wenxue Li[1], Dragan Maric[1], Farinaz Safavi[1], Iren Horkayne-Szakaly[4], Robert Jones[4], Michelle N. Stram[5], Marco Hefti[6], Rebecca D. Folkerth[5], and Avindra Nath[1]**

*** Presenting author**

**+ co-first author**

1 National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892, USA

2 Technology Access Program, NanoString Technologies, Seattle, WA 98109, USA

3 Department of Pathology, Uniformed Services University of the Health Sciences, Bethesda, MD 20814, USA

4 The Joint Pathology Center, Defense Health Agency, Silver Spring, MD 20910, USA

5 Office of Chief Medical Examiner, and Department of Forensic Medicine, New York University School of Medicine, New York, NY 10016, USA

6 Department of Pathology, University of Iowa Roy J. and Lucille A. Carver College of Medicine, Iowa City, IA 52242, USA

The neurological implications of coronavirus disease-19 (COVID-19), while still an active area of research, are becoming increasingly important to understand as more patients suffer lasting neurological complications from both acute and "long" COVID-19. Here we characterized the transcriptional signature across brain regions, as well as in specific neuronal compartments, for patients who died of COVID-19. We utilized digital spatial profiling (DSP) from NanoString Technologies (N= 3 COVID-19 patients and 2 controls) to study spatial transcriptomics as well as RNAscope from Advanced Cell Diagnostics (N = 9 COVID-19 patients and 10 controls) to study RNA in situ hybridization of post-mortem brain samples. RNA hybridization did not detect severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike or nucleocapsid transcripts in the brains of COVID-19 patients, however, mRNA transcripts of the complement genes C1QA and C3 were expressed in macrophages and endothelial cells of patient but not control brain samples. Analysis of the spatial transcriptomics data revealed distinct transcriptomic profiles for patient brain samples compared to controls in brain regions with high platelet endothelial cell adhesion molecule 1 (PECAM-1), a marker of vasculature, expression compared to regions with high protein tyrosine phosphatase, receptor type, C (CD45), a marker of leukocytes, expression. In patient samples, coronavirus signaling was most upregulated in the PECAM-1-rich brain regions, whereas transcripts of CAAT/enhancer-binding protein delta (CEBPD), a gene that regulates inflammation, was most upregulated in the CD45-rich brain regions compared to controls.

*(continue to page 110)*

*(continued from page 109)*

Analysis across brain regions indicates that wound healing, antigen presentation, and regulation of long non-coding RNA (lncRNA) pathways were significantly dysregulated in COVID-19 patients compared to control patients. APOD (encoding apolipoprotein D), CD74 (encoding an MHC-II protein), GSTT1 (encoding glutathione S-transferase theta 1) transcripts were the most specific upregulated transcripts in the brains of COVID-19 patients, compared to controls, across brain regions. Our study highlights the possibility that different neuronal compartments are differentially affected by COVID-19, and the therapeutic potential for targeting APOD, CD74, and GSTT1, and the proteins they encode, in the neuropathogenesis of COVID-19.

# Transcriptomic profiling of cardiac tissues from SARS-CoV-2 patients identifies DNA damage

**Arutha Arutha Kulasinghe**

University of Queensland

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is known to present with pulmonary and extra-pulmonary organ complications. In comparison with the 2009 pandemic (pH1N1), SARS-CoV-2 infection is likely to lead to more severe disease, with multi-organ effects, including cardiovascular disease. SARS-CoV-2 has been associated with acute and long-term cardiovascular disease, but the molecular changes that govern this remain unknown.

In this study, we investigated the landscape of cardiac tissues collected at rapid autopsy from SARS-CoV-2, pH1N1, and healthy control patients using targeted spatial transcriptomics approaches. The main outcomes for the study were to profile rapid autopsy samples collected from the COVID-19, pH1N1 and normal, non-viral deaths to determine transcriptional changes between the different cohorts. Although SARS-CoV-2 was not detected in cardiac tissue, host transcriptomics showed upregulation of genes associated with DNA damage and repair, heat shock, and M1-like macrophage infiltration in the cardiac tissues of COVID-19 patients. The DNA damage present in the SARS-CoV-2 patient samples, were further confirmed by γ-H2Ax immunohistochemistry. In comparison, pH1N1 showed upregulation of Interferon-stimulated genes (ISGs), in particular interferon and complement pathways, when compared with COVID-19 patients.

These data demonstrate the emergence of distinct transcriptomic profiles in cardiac tissues of SARS-CoV-2 and pH1N1 influenza infection supporting the need for a greater understanding of the effects on extra-pulmonary organs, including the cardiovascular system of COVID-19 patients, to delineate the immunopathobiology of SARS-CoV-2 infection, and long term impact on health.

# In silico saturation mutagenesis of SARS-CoV-2/ SARS-CoV spike and human ACE2 proteins

**Shaolei Teng[1], Adebiyi Sobitan[1], Vidhyanand Mahase[1], Qiaobin Yao[1], Dongxiao Liu[2], Qiyi Tang[2]**

1 Department of Biology, Howard University, 415 College St. NW, Washington, DC 20059
2 Howard University College of Medicine, 520 W Street NW, Washington, DC 20059.

The emergence of the pathogenic of pathogenic coronaviruses, including severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and severe acute respiratory syndrome (SARS-CoV) are serious threats to global health. The spike (S) glycoprotein of SARS-CoV-2 /SARS-CoV is responsible for the binding to the permissive cells, the first step of viral infection. The receptor-binding domain (RBD) of S protein directly interacts with the human angiotensin-converting enzyme 2 (ACE2) on the host cell's membrane. Therefore, S protein, RBD and ACE2 are critical for viral success to infection. In this study, we applied the computational saturation mutagenesis to mutate all residues in the S and ACE2 proteins to all other 19 amino acid types. We used structure-based energy calculations and sequence-based bioinformatics tools to quantify the systemic effects of missense mutations on the protein structure and function. A total of 18,354 mutations in SARS-CoV-2 spike protein were analyzed and we discovered that the majority of these mutations could destabilize the entire S protein and its RBD region. Specifically, the amino acid substitutions in SARS-CoV-2 RBD residue G431 can decrease the spike protein stability, but the mutations in its spatial residue S514 can make the spike and RBD more stable. Moreover, we showed that the most mutations in N-linked glycosylation sites, including N122 and N616, can increase the stability of the spike protein. In addition, we investigated 3,705 mutations in SARS-CoV-2 RBD and 11,324 mutations in human ACE2 and found that the mutations located in the interface of RBD-ACE2 complex can alter its binding affinity. Interestingly, SARS-CoV-2 neighbor residues G496 and F497 have different effects on RBD-ACE2 binding and ACE2 contact residues D355 and Y41 are critical for this protein-protein interaction. In addition, we found that L452R mutation can enhance the virus-host Interaction which may increase Omicron variant infectivity. We applied the similar approaches to investigate the effects of S mutations in SARS-CoV on protein stability and virus-receptor interaction. We found that target mutations of S protein amino acids generate similar effects on their stabilities between SARS-CoV and SARS-CoV-2. The analysis is critical for understanding the roles of coding mutations in SARS-CoV-2 spike on the viral pathogenesis of COVID-19. The findings provide potential target sites in the development of drugs and vaccines against COVID-19.

# A comprehensive and noninvasive prenatal screen to capture coding variation in fetal and maternal genomes

**Christopher W. Whelan[1,3,4], Harrison Brand[1,2,3,4,5], Michael Duyzend[1,3,7], John Lemanski[1], Monica Salani[1], Stephanie P. Hao[1,5], Isaac Wong[1], Elise Valkanas[1,3], Caroline Cusick[4], Lori Dobson[8], Courtney Studwell[8], Kathleen Gianforcaro[8], Stephanie Guseh[9], Benjamin Currall[1], Kathryn Gray[9], Michael Talkowski[1,2,3,4]**

1 Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA
2 Department of Neurology, Massachusetts General Hospital and Harvard Medical School
3 Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA
4 Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA
5 Pediatric Surgical Research Laboratory, Department of Surgery, Massachusetts General Hospital, Boston, MA
6 Department of Obstetrics and Gynecology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA
7 Department of Pediatrics, Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA
8 Center for Fetal Medicine and Reproductive Genetics, Brigham and Women's Hospital, Boston, MA
9 Division of Maternal-Fetal Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

Sequencing of cell free fetal DNA (cffDNA) from maternal blood has revolutionized prenatal diagnostics by providing a rapid screen for aneuploidies without invasive testing. Current cffDNA techniques are low resolution and mostly limited to screening for large chromosomal aberrations, while discovery of the majority of genetic variants that cause developmental disease requires invasive procedures (e.g. amniocentesis) followed by conventional genetic testing. In this study, we developed a novel approach to cffDNA sequencing and analysis based on high-coverage whole exome sequencing, which we call hrNIPS (High Resolution Non-Invasive Prenatal Screening). We designed methods for discovering and genotyping single nucleotide variants (SNVs), indels, and copy number variants in cffDNA data based on modeling the effects of the mixture of fetal and maternal cell free DNA on allele fractions, DNA fragment size distributions, and read depth. We have applied this method to over 50 pregnancies and show that it enables highly sensitive discovery of coding variation across the fetal exome at a low cost.

*(continue to page 114)*

*(continued from page 113)*

To demonstrate the feasibility of hrNIPS, we performed a pilot study on 51 pregnancies with gestational ages ranging from 9 - 40 weeks and generated an average of 190x exome-wide sequencing coverage. We benchmarked these methods against 10 samples with gold standard data from cord blood or amniocentesis DNA, and demonstrated high average sensitivity (89.2%) and precision (88.7%) for fetal SNV genotyping. We further assessed the clinical utility of hrNIPS across 14 fetal samples referred for invasive genetic testing as part of standard-of-care and demonstrated 100% concordance for discovery of diagnostic variants (n = 4). We also show that hrNIPS can provide added value in the form of maternal carrier screening, with high average sensitivity (96.2%) and precision (94.6%) when genotyping maternal SNVs, and find at least one reportable variant in over 50% of mothers in our study. In summary, we demonstrate the potential for hrNIPS to offer an accurate, non-invasive screen for diagnostic variants of relevance to maternal and prenatal health. Finally, hrNIPS provides a potential mechanism for integrating a fetal exome into the medical record at the earliest stages of development.

# Your single source for single-cell multiomics

## Introducing the new BD Rhapsody™ HT Xpress System

Be the first to see our new products in Suite 218.


**Gene expression analysis: Targeted or WTA**

**CITE-seq reagents for multiomic research**

**Immune repertoire profiling**

BD

# Other Biology Omics

# A spatially resolved, single-cell analysis of human olfactory cleft mucosa highlights the transcriptional dysregulation in sustentacular cells with SARS-CoV-2 viral load

**Jason W Reeves[1], Tyler Hether[1], Emily Killingbeck[1], Yan Liang[1], Mona Khan[2], Laura Van Gerven[3], Joseph Beechem[1], Peter Mombaerts[2]**

1 NanoString Technologies Inc., Seattle, WA, USA.
2 Max Planck Research Unit for Neurogenetics, Frankfurt, Germany.
3 Department of Neurosciences, Experimental Otorhinolaryngology, Rhinology Research, KU Leuven, Leuven, Belgium; Department of Otorhinolaryngology, Head and Neck Surgery, University Hospitals Leuven, Leuven, Belgium; Department of Microbiology, Immunology and Transplantation, Allergy and Clinical Immunology Research Unit, KU Leuven, Leuven, Belgium.
Electronic address:

Anosmia is a common symptom of COVID-19 and can persist well after viral clearance. This loss of smell is a result of the effects SARS-CoV-2 has on cell types that underly olfactory function. These cell types—which include sustentacular cells and olfactory sensory neurons—exist in a well-characterized spatial landscape called the olfactory mucosa, which consists of an archipelago of islands surrounded by respiratory mucosa. Our earlier work using the GeoMx® Digital Spatial Profiler characterized whole-transcriptome effects SARS-CoV-2 in the olfactory epithelium of a postmortem case at the spatial resolution of hundreds of cells and indicated that the transcriptional dysregulation of sustentacular cells was a likely driver of anosmia. This result raised the question of which genes in sustentacular cells are altered as a result of viral load. This inherently spatial single-cell question was addressed using the CosMx-TM Spatial Molecular Imager. Our panel consisted of 984 host targets and 9 probes for SARS-CoV-2, including Delta- and Omicron-specific variants. In total we measured 63,589,058 spatial transcripts in 401,233 cells. As a first step, we classified cells into the known cell types. We were able to classify cells in the surrounding lamina propria (Bowman's Gland, pericytes) using spatially agnostic and publicly available scRNA-seq data. A semi-supervised clustering algorithm allowed us to discern more nuanced cell types of the epithelium (olfactory vs respiratory horizontal basal cells) and to identify cell types that were not classified in the reference data (suprabasal cells).

*(continue to page 118)*

*(continued from page 117)*

Since SARS-CoV-2 infection results in degradation of host mRNAs, our approach was flexible enough to capture heavily infected cell types not adequately reflected in the reference (infected secretory, ciliated cells). Following classification, we then focused on the 725 sustentacular cells, grouped them into virus negative (535), low viral load (121), and high load (69), and found ~120 differentially expressed genes (DEGs). Differences between these groupings contain "classic" DEGs such as TMPRSS2 and genes related to inflammatory or myeloid signaling. As non-spatial studies of SARS-CoV-2 infection report on expression differences between individuals with different time-since-infection, we find a similar theme here and discuss how space may approximate time.

# An RNA exome panel used to enrich transcript variants using cDNA libraries



**Michael Bocek[2], Kristin Butcher[1], Yu Cai[1], Jean Challacombe[2], Derek Murphy[1], Esteban Toro[1]**

1 Twist Bioscience, Research & Development, South San Francisco, California, USA
2 Twist Bioscience, Research & Development Bioinformatics, South San Francisco, California, USA

Total RNA sequencing provides a relatively unbiased view of the transcriptional state of a population of cells. However, most total RNA-seq experiments must contend with a large number of reads that are not helpful for gene-expression analysis, including reads from highly abundant non-coding transcripts (like the 7SK RNA, or ribosomal RNA), intronic reads from pre-mRNA, or contaminating genomic DNA. Target enrichment provides a way to focus sequencing on the informative parts of the genome, allowing for more sensitive detection of low-abundance transcripts, or for profiling only specific genes of interest.

Here we present capture sequencing experiments using Twist's new RNA Exome panel, which uses a novel design strategy to specifically target every protein-coding isoform in Gencode v41 Basic. Although the design natively targets the transcriptome, our design strategy also places probes to minimize bias towards known isoforms, and allow for discovery of novel isoforms or fusion genes. We evaluate panel performance in expression quantification, showing that relative transcript abundances are preserved after hybrid capture. This allows for accurate and reproducible quantification of transcripts that are present across many orders of magnitude. Additionally, we show gains in sequencing efficiency from our targeted approach, and demonstrate the ability to capture novel structural variants such as RNA fusions common in cancers. Additionally, we discuss our bioinformatic approach to evaluating capture performance in RNA space, and discuss specific challenges in the analysis of RNA-seq experiments. In summary, we provide evidence that the Twist Targeted Enrichment for Gene Expression solution is an effective way to efficiently profile gene expression and detect gene fusions.

# Analyzing the molecular basis underlying anatomic and functional complexity of the mouse brain with MERSCOPE™



**Renchao Chen[1], Bin Wang[1], Cheng-Yi Chen[1], Nicolas Fernandez[1], Yuan Cai[1], Leiam Colbert[1], Jiang He[1]**

1 Vizgen, 61 Moulton St, Cambridge, Massachusetts, United States

Single-cell sequencing has provided numerous novel insights into cell heterogeneity and cell-type-specific adaptation of the nervous system. However, most sequencing-based techniques require cell dissociation and consequently lose the spatial information of the analyzed cells, which prevents directly connecting the cell types to specific anatomic and functional features. The rapid development of spatially resolved genomic assays enables molecular analysis in the tissue context, with the potential of revealing how single-cell gene activity orchestrates the structure and function of complex tissues like the nervous system. Here, we demonstrated the use of the MERSCOPETM Platform to generate a transcriptionally defined and spatially resolved single-cell mouse brain atlas. By performing multiplexed error-robust fluorescence in situ hybridization (MERFISH) assays with a 500-gene panel designed for cell typing, we obtained over one million cells with precise gene expression and spatial information across the

mouse brain. Clustering analysis of the gene expression data resolved all major cell populations as well as detailed neuron and non-neuron subtypes across different brain regions. Importantly, all these molecularly determined cell types were precisely mapped to their original locations. By assessing the relationship between molecular and anatomic features of identified cell types, we found transcriptionally defined cell types underlying the anatomic heterogeneity of different brain structures. Notably, integrating the spatial transcriptomic data obtained from MERSCOPE with single-cell RNA-seq/spatial genomic data provided novel insights into the molecular heterogeneity and functional complexity of mouse brain. Altogether, our work not only created a molecularly defined and spatially resolved mouse brain cell atlas, but also demonstrated the power of MERFISH measurements generated by the MERSCOPETM Platform in analyzing the molecular basis underlying the anatomic and functional complexity of the nervous system.

# AURORA: A simplified method for profiling protein translation

**Gene W. Yeo[2], Maya T. Kunkel[1], Alexander Shishkin[1], Stephanie C. Bruns[1], Kyle Stangline[1], Frederick Tan[2], Kylie A. Shen[1], Sergei A. Manakov[1], Osvaldo Rivera[1], Cassi M. Bruni[1], Karen B. Chapman[1], Eric L. Van Nostrand[3]**

1 Eclipse Bioinnovations Inc., San Diego, CA,
2 Department of Cellular and Molecular Medicine and Institute for Genomic Medicine, University of California San Diego, La Jolla, CA,
3 Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, One Baylor Plaza, Houston, TX

Widespread application of RNA-Seq has provided substantial insight into the role of gene expression in diverse biological processes. Quantitation of transcript levels by RNA-Seq, however, fails to capture regulation at the level of translation, which is responsible for a significant fraction of variation in protein abundance, and is linked to a wide array of human diseases. Here we present an approach to measure the transcriptome and translatome of any cell or tissue in an unbiased manner, in a singular workflow, without the need for specialized equipment or advanced techniques. Based on the well published eCLIP (enhanced crosslinking and immunoprecipitation) technology, AURORA involves enrichment of ribosome-associated RNAs to quantify protein translation by profiling ribosome-crosslinked RNA, while also providing corresponding RNA-Seq data. AURORA Total utilizes a simplified eCLIP protocol in which the cumbersome gel steps have been omitted from the workflow. AURORA Total offers ribosome occupancy and RNA-Seq data across the entire length of the transcript. AURORA 3' is an even more abbreviated eCLIP based method, utilized to count ribosome occupancy per gene, while providing 3' differential gene expression data. A comparison of AURORA and polysome profiling performed on Torin 1 treated cells reveals that both methodologies show a similar reduction in the ribosome-associated mRNAs containing TOP- and TOP-like motifs, while showing that no change in transcript levels was measured by RNA-Seq. The ability to quantitate ribosome load transcriptome wide using a simplified eCLIP methodology will enable translatome studies to be scaled more efficiently and applied more broadly in the context of research and in the development of therapeutics.

# CRISPR-Cas9–based repeat depletion for improved genotyping by sequencing of large genomes

**Massimo Delledonne[1], Luca Marcolungo[1], Luca De Antoni[1], Giulia Lopatriello[1], Elisa Bellucci[2], Gaia Cortinovis[2], Giulia Frascarelli[2], Laura Nanni[2], Elena Bitocchi[2], Kirstin Bett[3], Larissa Ramsay[3], David James Konkin[4], Roberto Papa[2], Marzia Rossato[1]**

1 Department of Biotechnology, University of Verona, Strada Le Grazie 15, 37134, Verona, Italy,
2 Department of Agricultural, Food and Environmental Sciences, Polytechnic University of Marche, via Brecce Bianche, 60131, Ancona, Italy
3 Department of Plant Sciences, University of Saskatchewan, 51 Campus Drive, Saskatoon, Saskatchewan S7N 5A8, Canada
4 National Research Council Canada, 110 Gymnasium Place, Saskatoon, Ontario S7N 0W9

High-throughput genotyping is increasingly required for large-scale diversity analysis, such as population genomics and GWA studies that combine genotypic and phenotypic characterization of large collections of wild and domesticated germplasm. Genotyping methods based on sequencing are progressively replacing traditional ones, but for large and highly repetitive genomes they suffer of high costs and of the coverage of many un-informative genomic regions.

We exploited the CRISPR-Cas9 system to deplete repetitive elements and concentrate sequencing data on meaningful coding and regulatory regions on lentil (Lens culinaris, 3.76Gb, 85% repeats). A custom set of 566.722 gRNAs - each with at least 25 recognition sites - targeting 2.9 Gbp of repeats was designed for 500bp-insert sequencing libraries. Repetitive regions overlapping annotated genes and putative regulatory portions identified based on ATAC-seq data were excluded. The CRISPR-Cas9-based depletion showed good efficiency, reducing coverage depth (down to 50%) on repeats, while improving the coverage on the functional fraction (2.7x). Thanks to such “repeat-to-coding” shift of sequencing data, the depletion allowed a net increase in the number of genotyped bases (up to 28x) in comparison to libraries not subjected to treatment. The method showed similar performances on low- and high-multiplexing levels and on different genotypes, including distinct cultivars as well as a close-related species (L. orientalis), thus demonstrating its potential for large genotyping studies. Overall, the CRISPR-Cas9-based repeat-depletion proved a cost-effective approach to focus sequencing data on functional genomic regions, thus providing an innovative opportunity to improve high-density and genome-wide genotyping in large and repetitive genomes.

# Detection of 5-hydroxymethylcytosine using a modified EM-seq method



**Daniel J. Evanich**, **V. K. Chaithanya Ponnaluri, Vaishnavi Panchapakesa, Ariel Erijman, Matthew A. Campbell, Nan Dai, Bradley W. Langhorst, Romualdas Vaisvila, Louise Williams**

New England Biolabs, Ipswich, MA 01938, USA

DNA methylation is an epigenetic regulator of gene expression with important functions in development and diseases such as cancer. Typically, the modified cytosines, 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC), are detected by sequencing Illumina libraries generated using an enzyme-based workflow called NEBNext® EM-seqÔ or bisulfite conversion. However, these methods cannot differentiate between 5mC and 5hmC. Identification of specific 5hmC sites is important as there is increasing interest in its role in regulating gene expression in embryonic stem cells and several neuronal cell types. Methods currently exist to enable discrimination of 5mC and 5hmC (for example, oxBS-seq and TAB-seq), however these are based on modifications of bisulfite sequencing, and suffer from reduced data quality due to fragmentation and loss of DNA. Here we describe an enzymatic method that enables specific detection of 5hmC.

5hmC is detected using two enzymatic steps. During the first step, 5hmCs are glucosylated, which protects them from subsequent deamination by APOBEC. In contrast, cytosines and 5mCs are deaminated resulting in their conversion to uracil and thymine, respectively. This conversion allows discrimination of 5hmC from cytosine and 5mC during Illumina sequencing. Additionally, subtracting 5hmC data from EM-seq data, which detects both 5mC and 5hmC, enables the precise location of individual 5mC and 5hmC sites.

5hmC data were generated for inputs of 0.1 ng to 200 ng DNA isolated from E14 mouse embryonic stem cells and human brain. The 5hmC libraries had similar characteristics to EM-seq libraries, including expected insert sizes due to intact DNA molecules, low duplication rates and minimal GC bias. T4147 phage DNA was used as an internal control, as all cytosines are 5-hydroxymethylated, with 98-99% of cytosines identified as 5hmC. 5mC and 5hmC levels were also profiled during E14 cell differentiation for a period of 10 days. 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. The ability to discriminate between 5mC and 5hmC will provide key insights into the role of these cytosine modifications in development and disease.

# Development and assessment of a robust workflow for tracking SARS-Cov-2 genomic variants and other pathogens in wastewater metagenome sequencing data



**Chunlin Xiao**[1]

1NCBI/NLM/NIH, 45 Center Drive, Bethesda, MD 20892

As SARS-Cov-2 virus continues to evolve and pose great threat to public health, early detection and constant surveillance of the viral genomic variants are essential for accurate assessment of community infection dynamics, and thereafter appropriate public health responses. Wastewater sequencing provides complementary solutions for tracking emergent viral genomic variants, particularly when routine clinical testing is not practicable due to resource limitations. However, proper analysis of wastewater metagenome sequencing data is a challenge due to the nature of mixture samples, low viral load, highly fragmented RNAs, and subsequent poor sequencing mapping rate and genome coverage. To date, different workflows have been reported, but they have not been systematically evaluated. Here, we report that by selecting four different BioProject wastewater metagenome datasets from NCBI SRA database, totaling 8,007 SRA runs (as of August 15, 2022), as our pilot study, we systematically evaluated the performances of 20 alignment-based (combinations of 4 aligners and 5 variant callers) and 3 de novo assembly-based analysis pipelines, and selected best-performed pipelines as our candidate production workflows based on a series of quality metrics. Both alignment-based and assembly-based SARS-Cov-2 genomic variants were combined and further evaluated for downstream analysis. For each sample, we observed that the percentages of reads being mapped to the de novo assembled contigs were significantly increased as compared to that with reads being mapped to the standard SARS-Cov-2 genome reference.

For those de novo assembled non-SARS-Cov-2 contigs, we used FCS-GX tool (https://github.com/ncbi/fcs) for taxonomy assignments, and found that the vast majority of the identified organisms were bacteria (approximately 91%) and virus (approximately 8%). Some were identified as human pathogens such as Streptococcus, Salmonella, E. coli, Clostridium, Cholera, Tuberculosis, and Gonorrhea etc. Moreover, a significant portion of well-supported contigs were not associated with any known organisms that were worth further investigation and exploration. This study illustrated that with carefully developed analysis workflows, wastewater metagenome sequencing could provide an opportunity for routinely monitoring of community-level SARS-Cov-2 genomic variants, but also could offer potentials for surveillance of regional microbial dynamics and infectious diseases.

# Dissecting transcription using temporally resolved high-throughput single-cell RNA sequencing

**Juliane Mayr[1], Christoph Ziegenhain1, Gert-Jan Hendriks[1], Anton Larsson[1], Michael Hagemann-Jensen[1], Rickard Sandberg[1]**

1 Department of Cell and Molecular Biology, Karolinska Institute, Sweden

Eukaryotic gene expression occurs in short periods of active transcription followed by periods of inactivity, a mechanism termed 'transcriptional bursting'. During each burst, the transcription machinery needs to be assembled at the respective genomic locus, directed by transcriptional co-factors. Recently, biomolecular condensates, membrane-less functional assemblies of biopolymers, have been proposed to facilitate the complex interplay of factors required for productive transcription. Technological advances in single-cell transcriptomics have now added temporal resolution through metabolic labeling which enables us to track short-term transcriptional output [1]. By increasing sensitivity and scalability of such methods, we can now capture direct transcriptional responses to cellular perturbations at the level of newly produced transcripts with single-molecule resolution. Here we employ emerging technologies to perturb the transcription machinery, either by targeting individual components or the entire transcription complex in the context of biomolecular condensates, to gain mechanistic insights into the functional basis of transcription. Understanding the role of individual transcription (co-)factors or multimolecular assemblies is necessary to delineate transcriptional plasticity and to map the regulatory networks underlying gene expression.

[1] Hendriks et al., Nature Communications, 2019

# Harnessing the power of evolution through pangenomes to understand biological traits



**Brenda Murdoch**,

University of Idaho

High quality reference genomes underpin all genomic research within a species. A wealth of scientific advancements, research tools, and discoveries have emerged based on establishment of reference genomes for livestock species, enabling major advances in the fields of animal science and genetics and benefitting livestock and associated industries for the benefit of society. A more complete understanding of how large and small genomic variants, such as those that are thought to exist in different genomes and have been selected or have evolved over time for specific environments and phenotypes, remains elusive. To address this, Pangenomes for multiple livestock species are being developed. The ability to have more than one reference genome for fine mapping causal variants or larger genomic regions associated with specific functional impacts will greatly enhance our ability to elucidate the underlying biological pathways. The long-term goal of this project is to provide genomic resources so that we can more completely represent the immense genetic diversity of this species. **This will increase our understanding of how various genomes result in different phenomes and empower scientists to more effectively predict the physiological outcomes from genomic variation towards the improvement of breeding strategies.** This type of genetic information constitutes the critical foundation necessary to discover the underlying mechanism of biological traits and inform future breeding strategies to support the sustainability of the supply of this global source of food and fiber.

# Introduction of a highly scalable and cost-effective proteomic immunoassay workflow



**Meghan Gillespie,** **Tim Desmet, Cole Walsh, Gregory Nakashian, Jack Pellegrini, Tom Howd, Sheila Dodge, Niall Lennon, Stacey Gabriel**

Broad Institute of MIT and Harvard, Genomics Platform, Cambridge, MA, USA

Multi-omics is a rapidly growing field that can help scientists to explore a "bigger picture" of biology by looking at both molecular and cellular aspects involved in disease. Proteomics specifically is important because of a protein's various functions, such as executing biological processes or indicating phenotypes in disease. It offers key benefits to improve drug discovery and development, which currently utilize genomics and transcriptomics as proxies for proteins. Proteomics also offers deeper insight into the dynamic changes in the human body relating to disease progression or in response to pharmacological interventions. Using Olink Proteomics' Explore 3072, immunoassays are multiplexed with thousands of specific biomarkers which can be used to screen plasma samples for a variety of proteins.

Proteomic methods were previously not sensitive or vigorous enough to measure the low abundant human plasma proteome. These methods were also very difficult to scale. A new proteomics method has been developed to measure the broad range of the plasma proteome, which is highly scalable and has reliable precision, sensitivity and specificity. This combines antibody-ligand binding assays with a DNA-based readout, which allows for this high throughput protein profiling while utilizing a low input material.

The Genomics Platform is working to build a high throughput proteomics lab. The process utilizes extremely low reaction volumes that would typically be processed by hand. To overcome these challenges we implemented low volume liquid handling instruments as well as integrated sample indexing and tracking using our in house LIMS systems. These systematic improvements allow us to utilize multiplexing across hundreds of protein biomarkers to be quantified simultaneously. This process is run in a high throughput lab environment while simultaneously analyzing data utilizing various types of sequencing.

Here the Genomics Platform presents a fully automated proteomic workflow which enables thousands of samples to be processed per month. This workflow provides a highly reproducible assay that is both highly scalable and cost effective, while requiring a low input material.

# Long-read sequencing reveals heritable large structural variants induced by CRISPR-Cas9

**Ida Höijer[1], Anastasia Emmanouilidou[1, 2], Lars Feuk[1], Ulf Gyllensten[1], Marcel den Hoed[1, 2], Adam Ameur[1]**

1 Science for Life Laboratory, Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden
2 The Beijer laboratory and Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden

Genome editing by CRISPR-Cas9 is an invaluable research tool that has potential to provide new treatments for genetic disorders. However, for clinical applications it is crucial to minimize the risk of adverse effects due to unintended CRISPR-induced mutations. To increase our understanding of unwanted CRISPR-Cas9 genome editing in vivo, we investigated the genomes of >1100 genome edited zebrafish larvae, juveniles and adults from two generations. Four guide RNAs (gRNAs) from previous functional studies were used to edit the genomes of fertilized zebrafish eggs at the 1-cell stage. Off-target sites were detected in vitro by nanopore sequencing (Nano-OTS). PacBio long-amplicon sequencing was used to investigate CRISPR-Cas9 outcomes at on- and off-target sites in edited zebrafish and their offspring. In founder larvae, on-target editing of the four gRNAs was 93-97% efficient, with high levels of mosaicism present in adult zebrafish. In vivo off-target editing was detected at three gRNA cleavage sites. Six percent of the CRISPR-Cas9 mutations corresponded to structural variants (SVs), including deletions <4.8 kb and insertions <1.4 kb. The F1 generation contained high levels of genome editing, with all alleles of 46 examined F1 juvenile fish affected by on-target mutations, including four cases of SVs. In addition, 26% of the juvenile F1 fish carried off-target mutations. We thus demonstrate that large SVs and off-target mutations can be introduced in vivo and segregate to the next generation.

Our results have important consequences for CRISPR-Cas9 ex vivo therapies, where cells from a patient are genetically enhanced outside of the body and reinjected as a treatment. Such treatments are now being developed for a wide range of inherited diseases, but also for cancer through modified immune cells. Long-read sequencing has the ability to detect unanticipated off-targets and SVs that may otherwise escape detection, and is therefore well-placed to become the gold-standard for genome editing verification. To streamline the CRISPR-Cas9 verification process, we are now exploring various targeted and whole-genome long-read sequencing strategies, and developing methods for computational analysis and reporting. We anticipate that our work will be an important step towards increasing the safety of future medical CRISPR-Cas9 applications.

# Multi-site assessment of methods for cell preservation upstream of single-cell RNA sequencing



**Fred Kolling,** Dartmouth College

Single cell RNA sequencing (scRNA-seq) is a revolutionary technique to identify cell types and their molecular phenotype in heterogeneous biological specimens. This high-resolution analysis is providing novel insights into the cellular architecture and homeostasis of tissues, as well as how these systems become dysregulated in disease and respond to treatment. scRNA-seq typically requires fresh, high quality single cell suspensions that are processed immediately to preserve their molecular profiles. This presents a challenge for samples with long preparation times, particularly patient samples that can spend hours in transit, and prevents collection at remote sites lacking the required instrumentation for sample processing. Recently, several commercial assays have been released that enable sample preservation at the time of collection either via fixation or cryopreservation, allowing for sample processing to occur weeks to months after the initial collection. At present, no comprehensive evaluation has been performed to assess how specific cell populations and their molecular profiles are impacted or how the performance differs between each commercially available preservation platform. The Association of Biomolecular Research Facilities' (ABRF) DNA Sequencing Research Group (DSRG) has undertaken a cross-platform, multi-site study to assess the performance and reproducibility of three platforms: 1) 10x Genomics Fixed RNA Profiling, 2) Parse Bioscience Evercode WT v2 and 3) Honeycomb Bio HIVE. Using total leukocytes from a single donor, we assessed the reproducibility of these three technologies and compared performance metrics including gene/transcript detection sensitivity, cell type discovery and annotation, and differential expression results. We also examine the ability of each method to retain the relative proportions of cell types from the original sample, using a 21-color flow cytometry dataset generated from the same donor in parallel as a reference. The improvements to preservation methods are changing the way research is conducted and our thorough investigation into the performance of each method provide a valuable resource to help scientists determine the most appropriate single cell preservation workflow given their sample collection logistics and laboratory infrastructure constraints.

# Multiplexed NGS-based QC tools for genome editing applications

Other Biology Omics

POSTER 412

**Joseph Mellor**[1], **Georgia Giannoukos**[2], **Maura Costello**[1], **Jack T. Leonard**[1]

1 seqWell, Inc., Beverly, MA
2 Editas Medicine, Cambridge MA

Genome editing with technologies such as CRISPR/Cas9 can be accompanied by off-target mutagenesis events that are critical to measure and characterize. For example, double-stranded DNA breaks at sites with near-matching target sequence occurs via nonhomologous end-joining, producing unwanted off-target mutations. Furthermore, DNA introduced to cells can insert at unpredictable locations, obliging a method to efficiently screen for all genome insertion sites in an unbiased manner. UDiTaS (Giannoukos et al., 2018) and GUIDE-Seq (Tsai et al., 2015) belong to a family of library prep methods aimed at detecting and quantifying these types of on- and off-target effects of CRISPR/Cas9 and other genome engineering enzymes.

In general, it is an effort-intensive, slow, and bias-prone process to generate libraries of genomic DNA using convential approaches such as fragmentation, size selection, end-repair and ligation prior to PCR. Library amplification is often followed by inefficient hybridization-based enrichment of sequences of interest to reduce sequencing cost.

Here, we describe the use of Tagify™ UMI Tn5 transposase-based reagents to circumvent these limitations by enabling streamlined, highly multiplexed, transposase-based workflows for putative target site or unbiased insertion site profiling. In this workflow, an initial tagging step is performed to add a full-length adapter carrying P5, an i5 index, a unique molecular identifier (UMI) upstream of the transposase-binding region and an Illumina-compatible sequencing primer. Next, a single or multiplexed set of user-customized PCR primers is used in combination with P5 primer to amplify genomic DNA between the custom primer(s) and the nearest transposition site. By titrating the ratio of UMI-tagging reagent to input genomic DNA, target or insertion site regions of different length ranges can be obtained. The UMIs carried on the full-length tagging adapters allow for correction of PCR bias introduced during library amplification and thereby provide an unparalleled comprehensive, quantitative read-out of both location and relative number of on-target and off-target sites in complex populations of cells. We further describe examples of the application of Tagify UMI Tn5 transposase-based Tagging Reagents and their associated workflow to characterize and quantify on-target versus off-target transgene insertions into the human genome.

# Nucleosome changes associated with cancer and NETosis-related diseases



**Theresa K Kelly[1], Kieran Zukas[1], Justin Cayford[1], Mark Eccleston[1]**

1 Volition America, 6086 Corte Del Cedro Carlsbad, CA 92011

Nucleosomes are the repeating unit of chromatin that contain important signals for proper genomic and transcription activity. Chromatin becomes decondensed in Cancer and NETosis related diseases resulting in nucleosome release into circulation as cell free (cfDNA). As a result, circulating nucleosome levels can be used as a surrogate for cfDNA. Furthermore, by retaining the histone-DNA complex that comprises a nucleosome one is able to measure a variety of genetic and epigenetic modifications that contain important information about the cells from which the nucleosomes were derived as well as the mechanisms by which they were released into circulation. We have previously shown that nucleosome levels are elevated in a variety of Cancers as well as patients with COVID and Sepsis. We have now assessed the correlation between nucleosome level and genomic integrity using our Nu.Q® assay and low coverage Oxford Nanopore sequencing. We found that cancer samples with high Nu.Q® levels have copy number alterations (e.g. Non-Hodgkin's Lymphoma sample with nucleosome level of 744 ng/ml displays amplifications in chromosomes 10, 11 and 18 and deletions in chromosomes 1, 8, 17, 18 and 23), whereas a sepsis sample with even higher nucleosome level (>1500 ng/ml) had no copy number alterations detected. Nucleosome modifications associated with Cancer have been studied for decades, but there is relatively less known about the nucleosome changes associated with NETosis and Sepsis. To study nucleosomal changes associated with NETosis we have developed an ex vivo NETosis induction system where we stimulate NETosis in whole blood. In contrast to studying NETosis in isolated Neutrophils, this synthetic sepsis system provides a model for understanding NETosis in a more natural microenvironment. Using both phorbol myristate acetate (PMA) and Calcium Ionophore we are able to increase nucleosome levels (>50X fold) and show that the different stimulants have different time courses of NETosis induction. Furthermore, we see changes in H3R8 Citrullination that are expected with NETosis. We will now apply our Nu.Q® Capture sequencing and mass spectrometry technologies to better understand the DNA and proteomic changes that occur during NETosis.

# Profiling genetic and epigenetic changes at read-level after cellular rejuvenation



**Dan Brudzewsky1, Diljeet Gill2, Aled Parry2, Fabio Puddu1, Casper K Lumby1, Nick Harding1, Joanna D Holbrook1, Páidí Creed1, Wolf Reik2,4,5,6**

1 Cambridge Epigenetix Ltd, Cambridge, United Kingdom
2 Epigenetics Programme, Babraham Institute, Cambridge, United Kingdom
3 Institution, Department, University of Cambridge, Cambridge, United Kingdom
4 Altos Labs, Cambridge Institute of Science, Cambridge, United Kingdom
5 Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom
6 Centre for Trophoblast Research, University of Cambridge, Cambridge, United Kingdom

DNA contains more information than the genetic alphabet A, C, G and T. Both genetics and epigenetics are fundamentally important to biology, and it is crucial to derive both the genetic and epigenetic dimension from DNA, in research and in the clinic. At the cellular level, ageing is associated with reduced function, altered gene expression and a perturbed epigenome. Recent work has demonstrated that the epigenome is rejuvenated by the maturation phase of iPSC reprogramming, which suggests complete reprogramming to iPSCs is not required to reverse ageing in somatic cells.

Here we apply the novel method "maturation phase transient reprogramming" (MPTR), where the factors required for reprogramming (OSKM) are expressed until this rejuvenation point and then withdrawn. To investigate genetic and epigenetic changes after transient reprogramming, we use the novel sequencing technology "CEGX 5-Letter seq". The current methods use next-generation sequencing (NGS) to elucidate epigenetic letters after differentially converting bases, dependent on their epigenetic modifications. However, discriminating epigenetic letters is accomplished by sacrificing the ability to distinguish all four genetic letters. CEGX 5-Letter seq addresses these problems by expanding the number of information states in NGS to sixteen. The expansion to sixteen states allows direct, digital and phased discrimination of genetic and epigenetic letters on the same read. Sixteen state encoding also enables suppression of artefacts leading to very high accuracy for both genetic and epigenetic sequencing.

*(continue to page 133)*

*(continued from page 132)*

Applying MPTR to dermal fibroblasts from middle-aged donors, we found that cells temporarily lose and then reacquire their fibroblast identity, possibly due to epigenetic memory retained at enhancers and/or persistent expression of some fibroblast genes. In addition, MPTR instigated a magnitude of rejuvenation that appeared substantially greater than that achieved in previous transient reprogramming protocols.

CEGX 5-Letter seq was performed on three experimental groups: Transiently reprogrammed (TR); Failed to transiently reprogram (FTR); and Negative controls (NC). The results show transient reprogramming induces genetic and epigenetic changes, including a large increase in allele specific methylation specific to TR samples. Further work is needed to understand the interplay between pre-existing or induced genetic variants, DNA methylation, and rejuvenation.

# Spatially resolved transcriptomics of astronaut skin biopsies reveals compartment-specific gene-expression changes during the Inspiration4 mission



**Eliah G. Overbey[1], Jiwoon Park[1], Sunny Narayanan[2], JangKeun Kim[1], Namita Damle[1], Deena Najjar[1], Krista Ryon[1], Matt MacKay[1], Evan E. Afshin[1], Justin Gurvitch[1], Richard Granstein[1], Sean Mullane[3], Briana M. Hudson[4], Aric Rininger[4], Sarah E. Church[4], Jaime Mateus[3], Christopher E. Mason[1]**

1 Weill Cornell Medicine, New York, NY, USA
2 Florida State University, Tallahassee, FL, USA
3 SpaceX, Hawthorne, CA, USA
4 NanoString® Technologies, Seattle, WA, USA

Recent developments in the private spaceflight sector have created new opportunities for omics research. Astronauts participating and leading private-sector missions can help us better understand the shifts in metabolic, immunologic, and biological homeostasis that are known to occur during spaceflight. In particular, spaceflight can cause epidermal hypersensitivity and interfere with epidermal proliferation and repair. To generate insight on the impact of short-duration spaceflight on the skin, we performed a 4mm skin punch biopsy on each of the four Inspiration4 crew members before (L-44 days) and after (R+1 day) spaceflight. We used the GeoMx® Digital Spatial Profiler (DSP) to quantitatively analyze spatially resolved gene expression levels. Specifically, we targeted four compartments across 95 regions of interest (ROIs) using the human whole transcriptome atlas (WTA, 18,676 genes) within the skin tissue. These layers were the (1) Outer Epidermis (OE), (2) Inner Epidermis (IE), (3) Outer Dermis (OD), and (4) Vasculature (VA).

Overall, we noticed 112 differentially expressed genes before vs after flight in the OE compartment, 20 in the IE, 138 in the OD, and 449 in the VA (padj < 0.1). Among these differentially expressed genes, we found loss of fibroblast and junction related genes in VA genes (i.e. DES, ACTA2, TLN1, TAGLN) and loss of keratin-related genes in the OD (KRT1 KRT5, KRT10, KRT14). Very few genes were differentially expressed across all ROI types (n < 10). However, gene set enrichment analysis (GSEA) revealed all compartments had an increase of KRAS signaling and inflammatory responses (q-value<0.01). We also quantified proportional changes in cell types between different skin compartments.

*(continue to page 135)*

*(continued from page 134)*

Melanocyte proportion decreased in the middle compartments (IE and OD) post-flight. Additionally, loss of T cells and pericytes were observed postflight in the VA compartment. These data represent the first-ever collection and analysis of skin biopsies from astronauts and have elucidated key changes that occur due to space-flight that we can use to design better countermeasures to protect future crews.

# The role of Shared Resources in facilitating human and environmental surveillance for SARS-CoV-2

**George Grills[1], Siôn Williams[1,3], Benjamin Currall[1], Natasha Schaefer Solle[1], Melinda Boone[1], Mark Sharkey[3], Braden Tierney[7], Jonathan Foox[7], Thomas Stone[1], Marissa Brooks[1], Corneliu Sologon[1], Elena Cortizas[1], Kristina Babler[4], Jiangnan Lyu[2], Zarnegarnia Yalda[2], Xue Yin[4], Krista Ryon[7], Bhavarth Shukla[6], Naresh Kumar[1], Dušica Vidović[1,5], Stephan Schürer[1,5], Christopher Mason[7], Helena Solo-Gabriele[4]**

1 Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, FL, USA
2 Clinical & Translational Science Institute, University of Miami Miller School of Medicine, Miami, FL, USA
3 Center for AIDS Research, University of Miami Miller School of Medicine, Miami, FL, USA
4 Environmental Engineering Laboratory, College of Engineering, University of Miami, Miami, FL, USA
5 Institute for Data Science and Computing, University of Miami, Miami, FL, USA
6 Department of Medicine, Division of Infectious Diseases, University of Miami Miller School of Medicine, Miami, FL, USA
7 Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY, USA

The Sylvester Comprehensive Cancer Center Shared Resources, working closely with other shared resources at the University of Miami (UM), helped establish and provide coordinated support for a multi-institutional study on environmental monitoring of SARS-CoV-2, the virus that causes COVID-19 disease, including surface, air, and wastewater-based sampling. This project provides a case study of how a diverse array of shared resources can work together to facilitate human and environmental surveillance for SARS-CoV-2. The study is a collaborative effort between researchers at UM and Weill Cornell Medicine. The shared resources involved in this project include a group of Sylvester Shared Resources, including the Behavioral and Community Based Research Shared Resource (BCSR), Biospecimen Shared Resource (BSSR), and Onco-Genomics Shared Resource (OGSR), along with the Miami Clinical and Translational Science Institute (CTSI) Biostatistics Collaboration and Consulting Core (BCCC), and the Miami Center for AIDS Research (CFAR) Laboratory Sciences Core. UM deployed an extensive human surveillance testing, tracking and tracing system to monitor students, faculty, and staff.

This study extended these efforts to encompass wastewater surveillance of SARS-CoV-2 from buildings on all the UM campuses, the city of Miami and surrounding county, public schools in the county, and UM-affiliated hospitals. The goals of this study are to generate, optimize, standardize, and compare SARS-CoV-2 human and wastewater surveillance with various sampling, processing, detection, and analysis techniques. The environmental viral surveillance data is integrated with community and hospital COVID-19 disease prevalence, with the aim of developing predictive models of local and regional level spread of the disease. The results from this effort are informing public health strategies on local and community levels and may serve as a model more broadly for other existing and emerging pathogens. We present here lessons learned, current results, and future directions, with a focus on the role and impact of the shared resources.

# Leveraging integrated multimodal single-cell data to drive the development of novel analytical algorithms



**Joseph Gans[1] , Daniel B. Burkhardt[2], Qing Liu[1], Kaylie Schneider[1], Laura Isacco[1], Cameron Reilly[1], Sakina Saif[1], Sunil Kuppasani[1], Hiyab Stefanos[1], Ben DeMeo[2], Peter Holderrieth[2], Sam Miller[2], Aqib Hasnain[2], Mrunali Manjrekar[2], Sophie Guo[2], Yash Raj[2], Allegra Lord[3], Mauricio Cortes[3], Aakash Kabbin[3], Ashwarya Chandler[2], Ethan French[2], Karen Gaffney[2], Jonathan M. Bloom[2], Scott Steelman[1], Chad Nusbaum[1]**

1 Cellarity, Technology Development, Somerville, Massachusetts, USA
2 Cellarity, Computational Data Science, Somerville, Massachusetts, USA
3 Cellarity, Biology, Somerville, Massachusetts, USA

Advances in single-cell genomics have in recent years enabled the measurement of multiple layers of genetic information across DNA, RNA, and protein. This increased access to information provides improved resolution into the drivers of health and disease at single cell resolution. These single-cell data types are the basis of our drug creation platform at Cellarity, where we leverage high dimensional data and machine learning to identify therapeutic interventions that reverse pathogenic cell states. For example, to strengthen the causal relationships across features, we can take advantage of orthogonal genetic information simultaneously within a cell (e.g. 10x Genomics Multiome ATAC-Seq + Gene expression). However, there are significant computational challenges that hinder optimal use of multimodal data for biological discovery. For example, single-cell data (multimodal data in particular) are sparse, often poorly annotated, and typically analyzed independently per modality. These challenges call for data-driven solutions to extract maximum value from multiple layers of genetic information and to elaborate on causal relationships at scale.

*(continue to page 138)*

*(continued from page 137)*

To spur development of novel algorithms addressing these issues, Cellarity has generated a multimodal data set of 300,000 CD34+ hematopoietic stem and progenitor cells derived from 4 healthy donors at 5 time-points. Each sample was measured using (i) Multiome Single-cell ATAC + RNA expression and (ii) Single-cell RNA + surface protein (CITE-Seq). To the best of our knowledge, this dataset is over 100x larger than any other publicly available multimodal time-course data. Cellarity organized the analysis by templating analyses and developing novel algorithms for cell annotation such that all data were analyzed in one week. This provides the basis for development of new tools for multimodal analysis. Cellarity has publicly released these data and is partnering with Open Problems in Single-Cell Analysis for use in their international competition to challenge competitors in predicting one modality from another (e.g. from DNA to RNA) at an unseen timepoint. In this presentation, we will discuss Cellarity's workflow for generating this data set as well as the computational methods used for multimodal analysis at scale.

# Mosaic brain aneuploidy of chromosome 1q originating preconception causes neurodevelopmental disease

**Katherine E. Miller[1,2], Jason Navarro[1], Jesse Westfall[1], Adithe Rivaldi[1], Jocelyn Lucyshyn[1], Sahib Sran[1], Mark Hester[1,2], Anthony R. Miller[1], Ryan Roberts[1,2], Maren Cam[1], Daniel Boue[1], Annapurna Poduri[3], Adam P. Ostendorf[1], Richard K. Wilson[1,2], Elaine R. Mardis[1,2], Erin Heinzen[4], Daniel Koboldt[1,2], Tracy A. Bedrosian[1, 2]**

1 Nationwide Children's Hospital; Columbus, OH, USA
2 The Ohio State University; Columbus, OH, USA
3 Harvard Medical School; Boston, MA, USA
4 The University of North Carolina at Chapel Hill; Chapel Hill, NC, USA

Acquired genetic mutations generate somatic mosaicism, a phenomenon in which genetically distinct cells exist within an individual. Depending on the developmental timing of the genetic alteration and its cell of origin, the resulting mosaicism may be distributed throughout the body or restricted to a particular tissue or cell type. Mosaicism is often assumed to be acquired as post-zygotic mutations, but few studies have tested this assumption. We performed 250x exome sequencing on diseased brain tissue from a cohort of 50 epilepsy patients and identified five patients with somatic copy number gains (1 or 2 extra copies) of the entire long arm of chromosome 1 (chr1q). The mutated cells had no representation in blood or buccal samples but were prominently observed in brain tissue. Four out of five patients had evidence of an extra haplotype (i.e., non-constitutional) inherited from mother for chr1q, suggesting that the copy number alteration occurred via meiotic nondisjunction. We utilized single cell DNA and RNA -sequencing and identified the chr1q gain to be present in ~10-30% of brain cells per patient, the majority of which were astrocytes. Our observations challenge the widely-held notion that somatic mosaicism is derived only from post-zygotic events and provides evidence for a distinct clinical entity caused by brain mosaicism arising from a pre-conception genetic alteration.

# The vaginal microbiome: Implications for treating vaginal disease



**Baris Ozdinc[1], Manoj Dadlani,[1] Kelly Moffat,[1] Dana M. Walsh[1]**

1 CosmosID

Vaginal microbiome communities can be categorized into community state types (CST) based on dominance by specific lactobacilli species. These CSTs are globally present and may indicate risk of vaginal infection. In this work we compare the vaginal microbiota of women from three different countries: Sweden (N = 45), China (N = 35) and two groups from the USA (N = 51 and N = 30). We imported this whole genome sequencing data from publicly available studies and performed a comparative microbiome taxonomic and functional analysis using the CosmosID-HUB. We found every CST within each group and the samples clustered according to CST rather than country, implying a broad applicability of the CST classification system. However, when we predicted functionality of the microbial communities, we found that they clustered by country rather than CST. This suggests that vaginal community functionality may be influenced by location, regardless of CST. One functional group consisted of samples with enriched abundance of genes coding for mucin degradation, which has been previously associated with pathogenic potential. The women in these group were from a mix of the three countries. These results suggest that, in a healthy state, CSTs are globally present, but their functionality may be influenced by location-specific factors. However, there may be a subset of functions with greater pathogenic potential that is not location-associated. These women may be more prone to vaginal infections and their identification may aid physicians in catching and treating vaginal disease early on.

# Progress in the identification and interpretation of noncanonical human translated sequences

**[1]Jonathan M. Mudge, [2]Jorge Ruiz-Orera, [3]John R. Prensner, [4]Eric W. Deutsch, [3]Irwin Jungreis [1]Adam Frankish, [4]Robert L. Moritz, [5]Sebastiaan van Heesch.**

1 European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, UK.
2 Cardiovascular and Metabolic Sciences, MDC, Berlin, Germany.
3 Broad Institute of MIT and Harvard, Cambridge, USA.
4 Institute for Systems Biology, Seattle, USA.
5 Princess Máxima Center for Pediatric Oncology, Utrecht, the Netherlands.

Extensive translation occurs outside of 'canonical' human protein and gene annotation catalogs. Especially, thousands of 'non-canonical' translations have been captured by Ribosome Profiling / Ribo-seq, which identifies translated ORFs by sequencing RNA undergoing ribosomal processing. Ensembl-GENCODE are spearheading collaborative, community-driven efforts to identify and classify Ribo-seq ORFs, and recently released an initial catalog of 7,264 translations into the reference annotation 'ecosystem'. Nonetheless, important questions remain on the functional interpretation of Ribo-seq ORFs; in particular, how many produce bona fide proteins? The assay does not demonstrate protein existence, and we have thus far classified only ~50 Ribo-seq ORFs as protein-coding by using evolutionary methods to distinguish protein-level constraint. Furthermore, our high stringency conventional mass spectrometry analysis finds little proteomics support for Ribo-seq ORF proteins. Are many translations simply molecular 'noise'? To investigate further, we have performed an extensive bespoke analysis of immunopeptidomics datasets. We observe that large numbers of Ribo-seq ORFs generate proteins that are presented by the HLA system within cell-lines or cancer samples. It could be that such proteins are essentially aberrant. Nonetheless, their implication in disease states (including in the work of others) has led to excitement in clinical science, considering – for example – their potential utility as biomarkers or therapeutic targets. Thus, we see value in producing a reference GENCODE annotation of 'non-normal' translation. Finally, we recognize that translation can instead have a purely regulatory function. Approximately half of our Ribo-seq ORFs are uORFs within protein-coding genes, and these sequence elements are also of emerging clinical interest. We will discuss the targeted improvement of our uORF catalog, including the development of a new evolutionary approach to infer their function.

# The NIAID (Gen)omics Program in support of SARS-CoV-2 research during the pandemic



**Inka Sastalla, Liliana Brown, Punam Mathur, Wiriya Rutvisuttinunt, Reed Shabman**

Office of Genomics and Advanced Technologies, Division of Microbiology and Infectious Diseases (NIAID), National Institutes of Health (NIH)

The National Institute of Allergy and Infectious Diseases (NIAID), through the Division of Microbiology and Infectious Diseases (DMID) has a comprehensive (Gen) omics Program and funds extramural research to generate and analyze genomic and transcriptomic datasets and to advance cutting-edge 'omics' technologies. These technologies enable studies to address a wide variety of biological questions related to pathogen evolution, fitness, virulence and interactions with its host and microbiome in order to support the development of novel therapeutics, diagnostics and vaccines. This Program is in line with NIAID's mandate to quickly launch a research response to newly emerging and reemerging diseases and biodefense Here, we provide a select overview of NIAID/DMID's pandemic response activities within the (Gen)omics Program. (1) the Genomic Centers for Infectious Diseases (GCIDs) apply and develop innovative sequencing technologies for high-throughput pathogen and host genomics. The GCIDs have made important contributions to outbreak investigations: they used viral genomics to understand super-spreader events in Massachusetts and Texas, and they support integrative SARS-CoV-2 genomic data analysis through the TERRA COVID-19 workspace. (2) The Structural Genomics Centers (SGCIDs) provide accurate and effective determination of 3D protein and protein-ligand structures; about 15% of SARS-CoV-2 protein structures in the Protein Data Bank have been contributed by the SGCIDs. (3) The Systems Biology for Infectious Diseases Program use large scale data types for computational modeling of host-pathogen interactions. They developed outbreak.info, an interactive website that aggregates biological and bioinformatic information into reports about SARS-CoV-2 variants. (4) The Bioinformatics Resource Centers (BRCs) provide a computational workbench for access to multi-omics data and bioinformatics analysis tools, including for SARS-CoV-2 genome assembly and annotation, phylogenetic trees of curated sequences, and sequence-variation and metadata-driven sequence analysis. During the SARS-CoV-2 pandemic NIAID/DMID leveraged these centers and other NIAID research programs in a concerted response to advance SARS-CoV-2 genomics and to drive discoveries in virus evolution, diagnostics, therapeutics, and vaccine development.

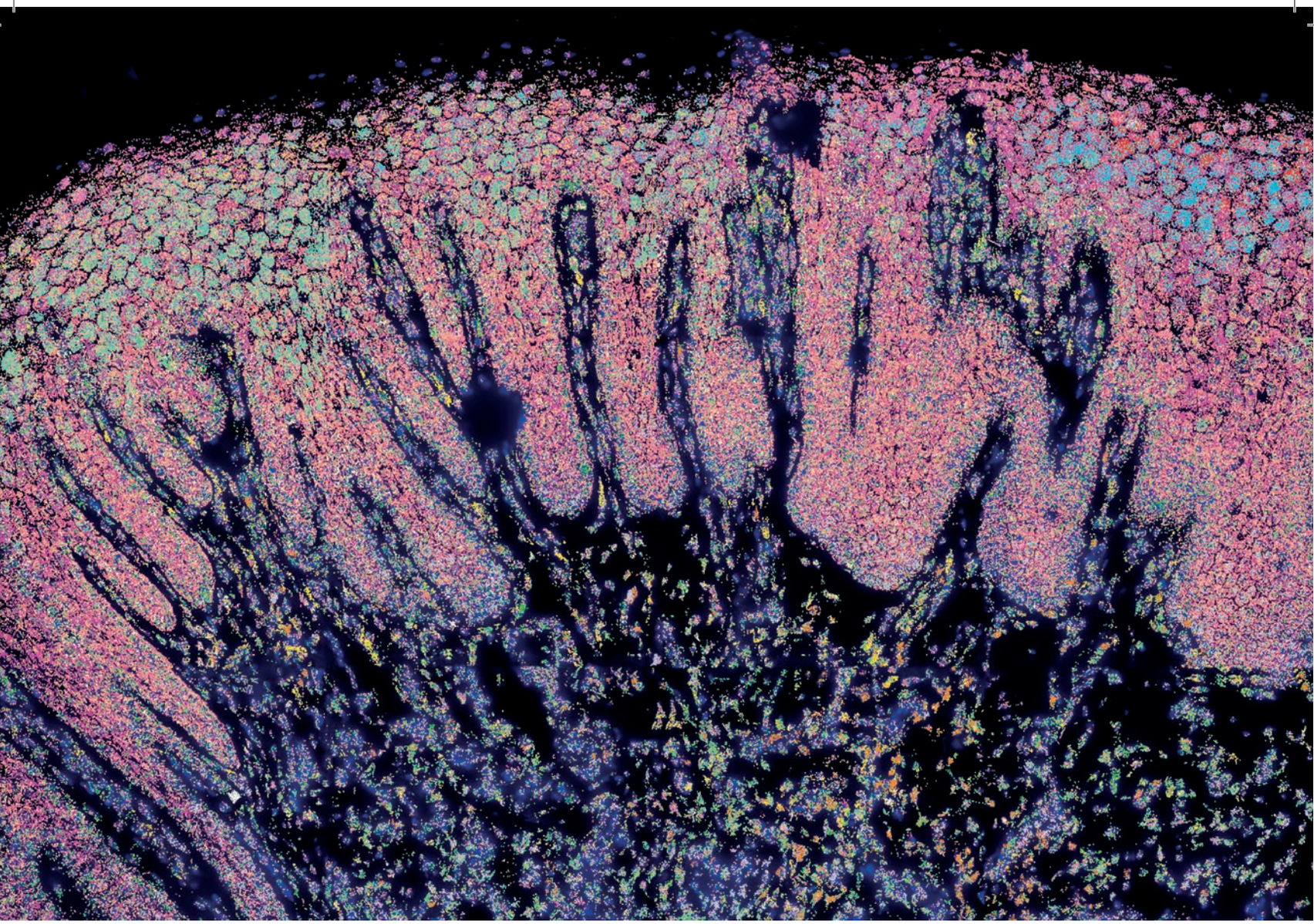
# Ruminant telomere-to-telomere assembly and analysis



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

1 USDA Agricultural Research Service, U.S. Meat Animal Research Center, Clay Center, NE, USA
2 University of Idaho, Dept. of Animal, Veterinary and Food Sciences, Moscow, ID, USA
3 University of Vermont, Dept. of Animal and Veterinary Sciences, Burlington, VT, USA
4 USDA Agricultural Research Service, Beltsville Agricultural Research Center, Beltsville, MD, USA

The Suborder Ruminantia has a high species richness in terms of extant organisms with 66 living genera in six Families. Ruminant species are more geographically dispersed and adapted to a wider variety of environments than existing primate species (except humans), providing a compelling resource for study of how environmental adaptation and domestication shape the genome. Chromosome number within the Suborder varies from 2n=30 to 2n=70, supporting analysis of how chromosomal fusions and fissions operate during speciation. The evolutionary resolution in the Ruminantia has large dynamic range spanning from species within Family Bovidae capable of interbreeding, to species with more than 2x difference in chromosome number. The Ruminant T2T project aims to generate complete genomes from species in each of the six main branches of the ruminant phylogeny to examine the chromosomal evolution and projection of the last common ancestor genome. Starting with available interbreed F1 crosses of domesticated species like cattle, sheep, and goat, we have generated initial assemblies that provide the first complete sex chromosome assemblies for these species. We have also generated T2T assemblies within the Bovidae for gaur, bison, and water buffalo. The assemblies and initial conclusions will be described, including the observation that rDNA arrays are not located on the short arms of acrocentric chromosomes as they are in humans, and that tangles in the sheep and cattle graphs are due to satellite repeats on the centromeric ends of telocentric chromosomes while rDNA arrays are essentially resolved.



# Focus on what matters

## Don't compromise on data quality. Every transcript counts.

Molecular Cartography™ data empowers scientists to generate high-impact, publication-ready results.

- Visualize gene expression patterns *in situ* with industry-leading sensitivity and specificity
- Create targeted gene panels specifically designed for your experiment
- Profile a broad range of tissue types, including FFPE and fresh frozen tissues
- Preserve your tissue sections post-analysis for additional study
- Send your samples to our experts for rapid, reliable results

**Join us daily in suite #213 to experience the power of Molecular Cartography data**

resolve biosciences

**Resolve Biosciences GmbH**
Creative Campus Monheim, Gebäude A03
Creative-Campus-Allee 12, 40789 Monheim am Rhein
Germany, (+49) 2173 2975-200

**Resolve Biosciences Inc.**
2890 Zanker Road, Suite 102
San Jose, CA 95134, United States of America
(+1) 408.944.5998

info@resolvebiosciences.com
www.resolvebiosciences.com

# Technology Development

# A cell-type enrichment strategy enhances the study of low-frequency somatic mosaic mutations

Technology Development

POSTER 500

**Jason B. Navarro[1], Jesse J. Westfall[1], Adithe C. Rivaldi[1], Richard K. Wilson[1], Elaine R. Mardis[1], Katherine E. Miller[1], and Tracy A. Bedrosian[1]**

1 Nationwide Children's Hospital, Institute for Genomic Medicine, Columbus, Ohio, USA

Mutations occurring in a small subpopulation of cells, such as non-inherited somatic mosaic mutations, can be challenging to capture via bulk whole exome sequencing (WES) due to low variant alle frequency (VAF). Somatic mutations are often restricted to particular cell lineages based on their developmental origin, so we hypothesized that enrichment of a specific cell type prior to WES can effectively increase the probability of finding low VAF somatic mutations. Here we investigate a fluorescence-activated nuclei sorting (FANS) approach to selectively enrich antibody-tagged nuclei isolated from frozen brain tissue before WES. Dubbed WESort (WES + cell sorting), this process is tunable to target a variety of cell types. In this work, we demonstrate the WESort workflow and show an example of it utilized with human brain tissue obtained from a patient with intractable focal epilepsy. A somatic mutation in a small percentage of cells resulting in a gain of chromosome 1q was discovered through initial exome sequencing, found primarily in the astrocyte population of affected tissue via single cell RNA-sequencing. However, not all astrocytes contained the chr1q gain, and astrocytes were found to be a very small percentage of tissue tested. We utilized WESort to enrich the astrocyte population via an antibody to PAX6, a nuclear transcription factor found in astrocytes, followed by sequencing of the PAX6+ and PAX6- populations of isolated nuclei. The enriched PAX6+ population contained a significantly higher percentage of nuclei containing the chr1q gain compared to the PAX6- population, as well as a higher percentage than found in initial non-enriched sample sequencing, allowing for much greater sequencing depth of the mutation of interest. The method and workflow of WESort is also adaptable to work with other downstream next-generation sequencing methods such as ATAC-seq, ChIP-seq, and single cell RNA-seq.

# A new platform using deep-learning models to perform multi-dimensional morphology analysis for biological discoveries

POSTER 501

**Stephane C. Boutet,** **Chassidy Johnson, Hou-Pu Chou, Rudy Hofmeister, Nianzhen Li, Kevin B. Jacobs, Kiran Saini, Senzeyu Zhang, Ryan Carelli, Amy Wong-Thai, Vivian Lu, Andreja Jovic, Anastasia Mavropoulos, Mahyar Salek, Maddison Masaeli**

Deepcell, Menlo Park, California, USA

Morphology is a fundamental cell property associated with cell identity, state, function and disease. Current methods to quantify morphology lack high discriminative power and are generally destructive to cells, limiting data that can be extracted from samples. Morphology as a multi-dimensional readout, 'morpholomics', can provide a more comprehensive understanding of cell biology, particularly when integrated with other multi-omics approaches. Accordingly, we applied artificial intelligence (AI) and computer vision to characterize cells following real-time deep learning interpretation of high-content morphology data and collect label-free, unperturbed cells for further characterization and integration with other cell analysis modalities.

Single-cell suspensions, including isolated from dissociated tissues, are loaded onto the instrument and imaged at high speed using 63X magnification, yielding high resolution (0.16 µm/pixel) images. As the label-free cells flow through the microfluidic channel, AI models analyze the brightfield images in real-time. Cell types and/or subpopulations of interest can be retrieved from one of six wells for further multi-omic and functional analysis enabling integration of mopholomics with other -omics modalities.

AI embeddings are reproducible, quantitative, and multi-dimensional descriptions that distinguish individual cells from one another. However, AI embeddings are not conventionally interpretable to biologists. To aid the translation to biological significance, we combined machine learning with computer vision to develop a generalizable AI model, termed 'General Human Model' (GHM), for multi-dimensional morphology analysis on a broad range of human cells. The GHM training and validation datasets consisted of 1.18 million and 4.05 million cell images, respectively, from multiple cell lines to capture diverse cell morphologies and cover the spectrum of biological diversity.

*(continue to page 148)*

*(continued from page 147)*

The Deepcell Cloud allows users to explore these AI embeddings through interactive UMAPs, population analytics, and image visualizations. Using these features, investigators can define and save cell populations of interest, and reproducibly use them for subsequent analysis or retrieval. Together, the Deepcell platform and Cloud can help usher quantitative morphology as a new "omic" modality by revealing morphology-based information critical for biological discoveries.

# A perspective on a molecular approach on newborn screening in the Netherlands



**Marcel R. Nelen[1], D. Westra[2] , M.R. Heinder-Fokkema[3], B. Raddatz[3], E. Voorhoeve[4], M.B.G. Kieviet[3], FJ. Spronsen[3]**

1 UMC Utrecht, the Netherlands.
2 R,adboudumc, The Netherlans
3 UMC Groningen, the Netherlands
4 RIVM, The Netherlands

Newborn blood screening (NBS) aims to identify newborns in which immediate treatment saves life or improves clinical outcome. NBS analyses dried blood spots (DBS) for abnormal metabolite concentrations. The use of next generation sequencing (NGS) techniques aims to improve current NBS. NGS-first techniques for NBS would allow detection of diseases that miss a detectable biochemical footprint and thus may antedate diagnosis of diseases. Pilot studies explored possibilities and challenges of genomic analysis in NBS. The results of these studies show a promising future. The actual step of applying NGS analysis in a routine setting remains a challenge. Our project explores how to make this next step possible taking inborn errors of metabolism (IEM) as example. To identify the challenges ahead to scale up routine NGS clinical workflows to a capacity that would fit the numbers of newborns (±180.000) in the Netherlands we have tested 48 both blinded NBS positive NBS cases together with 48 random negative dried bloodspots. Samples were tested with a targeted panel approach (coding sequence of 100 genes, Agilent), a routine WES (virtual gene panel, Twist), and a WGS based approach. Data analysis was done using in-house developed bioinformatic pipelines used routinely in clinical genetic testing workflows. The main purpose was to identify the burden it would present on the interpretation aspect of the NBS-workflow.

For the 50 NBS positive bloodspots, a summary of the targeted / WES / WGS outcome. In 29 cases there is a VUS, likely benign, or benign variant present after the first filteringstep (exonic and splice sites). In 38 cases either two causal variants were detected (30) or likely pathogenic (LP) variants in combination with a VUS (8, all in the same gene). In 4 cases causal variant were outside routine filtersettings or a homozygous VUS was present. In 4 cases causal variants were identified but were carriers of additional (L)P variants in other genes In the remaining 2 cases only one causal variant was detected (even with no filtering) and one case was identified as a carrier. The 48 negative bloodspots detected 7 carriers, 21 had a VUS, LB of B variant after the first filterstep.

*(continue to page 150)*

*(continued from page 149)*

Next approach was to examine 5000 anonymized parents of ID-families send in for routine WES analysis. The same IEM gene panel was applied to the available WES-data. In summary, we see 30 individuals with (L)P that fits the mode of inheritance of diseases in this gene-panel. If this is the case, they were either in cis or literature pointed towards very mild symptoms, or they were characterized as asymptomatic variants.

In conclusion, all three approaches provided acceptable test results and were in concordance with each other. The burden on the routine interpretation of a molecular based NBS is likely limited. However, one has to have a close look at all variants that would provide a positive NBS test result to exclude in cis variants or variants that provide mild or even asymptomatic symptoms.

# Adaptive sequencing of insect genomes to enrich for ecologically relevant endosymbionts

**Sara Goodwin[1], Jonathan H. Badger[2], Rosanna Giordano[3], Aleksey Zimin[4], Robert Wappel[1], W. Richard McCombie[1].**

1 Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA
2 National Cancer Institute, Laboratory of Integrative Cancer Immunology, Bethesda, MD, USA
3 Florida International University, Institute of Environment, Miami, FL, USA
4 Johns Hopkins University, Dept. of Biomedical Engineering, Baltimore, MD

Bacterial symbionts of insects are important organisms that can alter an insect's phenotype in many diverse ways. This can include the provision of essential nutrients, protection from natural enemies, or altering the host's method of reproduction. Furthermore, many insects exhibit co-infection with multiple symbionts which can further affect their phenotypes. Currently, two methods are used to sequence the genomes and plasmids of symbionts; whole genome sequencing on an individual/population capturing the bacterial genome along with the host's, or enrichment, via amplification or capture to isolate the symbiont's genome from the host's. However, these methods are limited when dealing with lower abundance or poorly characterized symbionts. Here we describe a method to selectively sequence insect symbionts without a priori knowledge of their composition using the soybean aphid as an example.

DNA from soybean aphids was extracted via phenol-chloroform and prepared for Oxford Nanopore sequencing. A fasta file of the aphid genome without its symbionts and mitochondria was prepared and used along with adaptive sampling on an Oxford Nanopore GridION to reject reads mapping to the aphid genome while maintaining reads mapping to the symbionts, their plasmids, and the mitochondria. More than 9M reads were generated which were then assembled with the metaFlye assembler, designed for long read data from a mixture of taxa. The resulting contigs were subsequently classified using Kraken 2 with the database created for the JAMS pipeline, which includes bacteria, viruses, and some eukaryotes (insects, human, mouse).

*(continue to page 152)*

*(continued from page 151)*

Approximately 50% of the reads aligned to the host genome, 36% mapped to Buchnera, 0.75% to Wolbachia, 0.62% and 0.26% to the Buchnera pTrp and pLeu plasmids respectively, and 0.1% to the mitochondria genome. Assembly of the symbiont's genomes yielded single contigs representing their entire genomes. In the case of the plasmids and mitochondria, single reads were found to span the entire genomes. This method will be an important tool for surveying symbiont populations for many insects of agricultural and medical importance as well as infected tissues of other eukaryotes.

# An optimized linked-read method empowers short-read sequencers to phase 2 to 200 kilobase DNA targets



**Ivan Garcia-Bassets[1], Veronika Mikhaylova[1], Madison Rzepka[1], Tetsuya Kawamura[1], Long Pham[1], Yu Xia[1], Peter L. Chang[1], Likun Yao[2], Adrian Perez-Agustin3, Sara Pagans[3], Christian Boles[4], Ming Lei5, Yong Wang[5], Zhoutao Chen[1]**

1 Universal Sequencing Technology Corp., Carlsbad, CA 92011, USA
2 Cellular & Molecular Medicine Department, University of California, San Diego, La Jolla, CA 92093 USA
3 Medical Sciences Department, School of Medicine, University of Girona, Girona, Spain
4 Sage Sciences Inc., Beverly, MA 01915, USA
5 Universal Sequencing Technology Corp., Canton, MA 02021, USA

Humans carry two copies of each homologous chromosome. But the two sets of chromosomal copies have millions of allelic differences. Phasing refers to resolving these differences by chromosomal copy, which is critical for many clinical and research applications (e.g., genomic testing and predictive genomics). Transposase enzyme-linked long-read sequencing (TELL-SeqTM) is a linked-read library prep method that empowers short-read sequencers to phase whole genomes. However, it is more cost-effective for many applications to phase a targeted region covering a few allelic differences (haplotype) than a whole genome. A targeted approach requires target DNA enrichment prior to phasing. The downside of this strategy is that enriched targets are very small with an ultra-low complexity compared to a whole genome, which traditionally is the desirable TELL-Seq sample. Here, we describe and validate an optimized TELL-Seq method that enables accurate phasing of 2-200 kilobase (kb) genomic and transcriptomic targets obtained with different DNA-enrichment methods. Only a few picograms of starting material are required for successful phasing (20 pg or less), and we estimate that a panel of up to 250 targets (if they map to different genomic or transcriptomic regions) could be pooled in a single TELL-Seq reaction. In support of the accuracy of the optimized method, we show complete haplotype reconstruction of five loci with clinical significance (SCN10A, PIK3CA, MLH2, BRCA1, and BRCA2).

*(continue to page 154)*

*(continued from page 153)*

DNA was obtained from cell cultures or whole blood using different methods, including high-salt DNA extraction, no extraction (cell lysis), and cDNA. The targets were enriched according to DNA length. We conducted PCR amplification for the smallest DNA targets (from 2 kb to 13 kb), Cas9-mediated protection followed by exonuclease treatment for the medium-size DNA targets (from 20 kb to 40 kb), and CRISPR/Cas9-mediated excision followed by pulse-field electrophoresis for the largest DNA targets (from 180 kb to 200 kb). To aid the phasing analyses, we developed a pipeline that conveniently outputs TELL-Seq phased alleles on the Integrative View Genome (IGV). Together, our results reveal high phasing accuracy of clinically relevant targets ranging from 2 kb to 200 kb using TELL-Seq without the need for a long-read sequencing platform.

# Application of ONT long-read sequencing to confirm microdeletions and microduplications in a clinical setting

Technology Development

POSTER 505

**Stephanie Greer[1], Jacquelin Botello[1], Donna Hongo[1], Brynn Levy[2], Premal Shah[1], Matthew Rabinowitz[1,3], Kate Im[1], Akash Kumar[1]**

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

Genomic copy number variants (CNVs) underpin many neurocognitive disorders, including intellectual disability, developmental delay, and autism spectrum disorders. Long read sequencing with Oxford Nanopore is a promising technology for clinical testing to identify causal variants in disease in a single laboratory assay, with the ability to test multiple genomic features, including sequence variants and methylation, and address a variety of existing challenges, such as read alignment in difficult-to-analyze genomic regions. Here, we successfully validated the Oxford Nanopore long read technology in a clinical assay to confirm the presence of CNVs initially detected using short-read WGS.

To validate our approach, we sequenced 24 Coriell samples and 5 blood samples with known microdeletions/duplications using an ONT GridION sequencer with adaptive sampling. Adaptive sampling enables the real-time selection of DNA molecules in user-specified genomic target regions, allowing us to achieve adequate on-target depth using a single flowcell. Each sample harbored at least one known CNV detected previously by microarray and short-read WGS analysis, such that we assayed a total of 35 unique CNVs ranging in size from 40kb to 155Mb. Further, the robustness of our assay was tested by assessing intra-run and inter-run concordance of samples. We developed a bioinformatic pipeline, implemented in Nextflow and executed on AWS, to assess the presence or absence of suspected CNVs using normalized read depth. A duplication or deletion was confirmed if the read depth across the suspected variant region was at least three standard deviations above or below, respectively, the mean depth of five genomic control regions.

*(continue to page 156)*

*(continued from page 155)*

Across 49 MinION flow cells, we achieved an average on-target mean depth of 9.4X (range: 2.7X - 22.7X). The average on-target read length across flow cells was 5316bp. Using a read depth analysis of our on-target data, we successfully confirmed all 55 of the expected CNVs across 49 samples (including replicates). Our results represent an initial clinical application of ONT long read sequencing to confirm CNVs. Overall, targeted ONT long read sequencing can enable improved variant discovery in difficult-to-analyze genomic regions, simultaneous methylation and sequence analysis, and ultimately an improved rate of clinical diagnosis.

# Automated, ultra-high-molecular-weight DNA library preparation for Oxford Nanopore Technologies platforms on the Miroculus Canvas system

Technology Development

POSTER 506

**Spyros Oikonomopoulos[1,2], Julio Serna-Vazquez[1,2], Mina Lotfi[1,2], Yen-Yen Wang[1,2], James C Engert[4], Cheng-Chang Lee[5], Ankita Chavan[5], Sridhar Mandhali[5], Jiannis Ragoussis[1,2,3]**

1 McGill Genome Centre, McGill University, Montreal, QC, Canada;
2 Department of Human Genetics, McGill University, Montreal, QC, Canada;
3 Dept Bioengineering, McGill University, Montreal, QC, Canada;
4 Division of Experimental Medicine, Department of Medicine, McGill University, Montreal, QC, Canada;
5 Miroculus, Inc., San Francisco, CA, USA.

Long-read sequencing has emerged as an important tool for the characterization of clinical samples. Within the long-read sequencing realm there is utility in obtaining ultra-long sequencing data to identify large structural variants and decoding complex disease mechanisms.

Manual preparation of such libraries can be complicated and requires a demanding skill set by the operator who is challenged to maintain the integrity and length of ultra-high molecular weight DNA throughout protocol steps. To help streamline this process and broaden the reach of this application in the clinical setting, we used the Miro Canvas to develop sample preparation methods for ultra-long read DNA sequencing.

Modifications of some Oxford Nanopore Technologies (ONT) 1D ligation kit protocol steps in conjunction with microfluidic processes in Miro Canvas, that minimize the shearing effects pipetting has on long DNA fragments, allowed us to robustly prepare libraries for ultra-high long read sequencing, ensuring N50 scores of as long as 77kb in obtained datasets

*(continue to page 158)*

*(continued from page 157)*

All of the sample processing steps including enzyme additions, incubations, and magnetic bead-based clean ups are done on the Miroculus instrument. The recovered % of library in relation to the input amount of 1ug of DNA ranged between 15-40% across the different clinical samples. In all experiments, the output was sequencing ready and was loaded on an ONT nanopore flowcell for immediate sequencing.

To establish this protocol, we used both reference DNA (NA12878) as well as samples from patients with familial dilated cardiomyopathy (DCM). Truncating variants in the TTN gene are the commonest cause of DCM. We are working on cases without a genetic diagnosis and we have samples with evidence of an inherited condition derived from affected, and a few unaffected, family members.

We report here that this technology matches manual techniques in terms of turnaround time (~3 hours of library preparation time), reproducibility and data quality. This new workflow can lead to better data collection for clinically relevant human genomes due to minimal manual intervention in the library preparation process. As an automation platform, it offers the user a hands-off library preparation method especially in cases when there is a need to sequence quickly on a nanopore sequencer.

# Chromium Fixed RNA Profiling reveals single-cell data with comparable biological complexity from FFPE blocks and their matching fresh tissues

Technology Development

POSTER 507

**Ryan Stott**,
10x Genomics

Formalin-fixed, paraffin-embedded (FFPE) tissue blocks are one of the most commonly generated and archived sample types, with enormous potential as a resource for future transcriptomic studies. However, FFPE preservation results in the degradation of RNA which hinders reverse-transcription-based transcriptome profiling. The new Chromium Single Cell Fixed RNA Profiling assay from 10x Genomics uses in situ hybridization of probes to RNA to assess whole transcriptome expression of FFPE samples, allowing the detection of low-expressing genes with high sensitivity.

To assess data quality, especially the level of biological complexity that can be obtained from FFPE blocks, we compared single cell sequencing data from FFPE blocks to data from corresponding fresh tissue samples using the Chromium Fixed RNA Profiling assay. We tested multiple mouse tissues including brain, liver, kidney, and thymus. Half of the fresh organs were immediately fixed and dissociated, while the other half were made into FFPE blocks, followed by sectioning and dissociation. Both sample types were hybridized with the mouse whole transcriptome panel, followed by probe ligation, barcoding, and sequencing. Comparable cell proportions and clustering data were observed between the FFPE blocks and their matching fresh tissue pairs. This finding demonstrates the robustness of the assay in supporting accurate biological discoveries.

Our results demonstrate the capability of the new Chromium Fixed RNA Profiling assay to unlock the large number of archived FFPE samples for single cell transcriptomic analysis. This approach is a powerful tool for clinical and academic researchers as it enables gaining unprecedented biological insights from archival FFPE blocks.

# Clinical validation of end-to-end whole genome sequencing and interpretation applied to both panel-based and diagnostic genome testing

Technology Development

POSTER 508

**Katherine A. Lafferty[1], Diana M. Toledo[1], Areesha Salman[1], Kathleen C. Lewis[1], Edyta B. Malolepsza[1], Katie L. Larkin[1], Vanisha Mistry[2], Erwin Frise[2], Weimin He[2], Jeanette McCarthy[2], Stacey B. Gabriel[1], Martin G. Reese[2], Heidi L. Rehm[3,1], Niall J. Lennon[1]**

1 The Broad Institute of MIT and Harvard, Cambridge, MA, USA
2 Fabric Genomics, Oakland, CA, USA
3 Massachusetts General Hospital, Boston, MA, USA

The Clinical Research Sequencing Platform (CRSP) at the Broad Institute is a CLIA-licensed and CAP-accredited clinical laboratory that provides clinical whole genome sequencing to partner laboratories and clinical research programs. Currently, customers are provided with raw data or secondary analysis VCFs for use in their external pipelines. Here, we describe a plan to support and validate panel-based or indication-based interpretation, enabling launch of an end-to-end whole genome service with interpretation.

To enable high-throughput interpretation services, CRSP has partnered with Fabric Genomics. Panel-based testing is facilitated by Fabric's AI Classification Engine (ACE), assisting in the auto-application of certain types of evidence according to ACMG/AMP criteria, highlighting potentially pathogenic variants for further review. The ACMG v3.1 secondary findings list of 78 genes was used for ACE validation. Indication-based testing is facilitated by Fabric's GEM algorithm, which prioritizes all variants in a genome by their likelihood of causing a given phenotype.

A cohort of clinical and reference samples with variants of interest was used to validate this service. The variant spectrum includes single nucleotide variations, small insertions and deletions, and larger copy number variants. Between 20-25 samples of each variant type were identified, using unique samples for ACE and GEM algorithms, resulting in ~136 validation samples.

*(continue to page 161)*

*(continued from page 160)*

We present here the results of the validation study. The validation assessed whether the variant of interest was detected and prioritized for review as expected and where the variant ranked within the GEM algorithm. To date, 43 total cases with small variants were run through an ACE panel, and 30 total WGS were run through GEM variant prioritization. All 43/43 ACE cases auto-classified the variant of interest as a VUS or above, requiring additional review. 29/30 (96.7%) WGS cases run through GEM plus systematic application of secondary filters (Phevor ranking, ClinVar, MAF) identified the variant of interest. We also reviewed the manual variant analysis steps for each classified variant to assess accuracy of the manual portion of the classification. The algorithm was also assessed for repeatability and the platform was assessed for security. The full outcome of this validation will be presented.

# Combinatorial high-throughput genetic screening to reprogram the fate and function of therapeutic T cells



**Grace Zheng, Brendan Galvin, Angela Boroughs, Ashley Cass, Luke Cassereau, Joe Choe, Aaron Cooper, David DeTomaso, Shen Dong, Fernando Fisher, Ryan Fong, John Gagnon, Levi Gray-Rupp, Angela Henkensiefken, Keith Joho, Sahil Joshi, Jessica Kotov, Sofia Kyriazopoulou Panagiotopoulou, Andrea Liu, Christopher Murriel, Dina Polyak, Gavin Shavey, Rene Sit, Josephine Susanto, Michelle Tan, Colin Tang, Sam Williams, Kristina Vucci, Sophie Xu, Tarjei Mikkelsen, W. Nicholas Haining**

Arsenal Biosciences

CAR T cell therapy has emerged as a critical tool in the treatment of hematologic cancers, but successes in treating solid tumors with CAR-T cells are rare. Clinical efficacy in solid tumors will require engineering T cells with increased therapeutic potency and high tumor specificity. To this end, ArsenalBio's Integrated Circuit couples logic-gated cytotoxic activity with additional genetic engineering features that enhance proliferation, resist exhaustion, overcome tumor microenvironment-mediated suppression, and improve anti-tumor efficacy.

To efficiently and effectively survey the maximal number of features that enhance T cell fate and function, we developed a single-cell screening platform that uses pooled and arrayed in vitro and in vivo screens to identify and characterize potential modulators of T cell function. Pooled screens enable comparison of a larger number of perturbations whereas arrayed screens allow for more complex functional assessment. In all screens, we use single-cell sequencing to comprehensively assess cell states and as a map to bridge functional readouts to genetic perturbations. For each screen, three modalities were read from each cell: gene expression; cell surface protein expression; a perturbation barcode. Different approaches are used to capture the perturbation barcode depending on the screen format: 1) direct capture of guide RNAs; 2) lipid-modified oligos  3) barcodes incorporated within 5' UTRs of the perturbations. This multimodal 5' single-cell screening platform used across pooled, arrayed, and in vivo formats combines the best features of numerous technologies into a streamlined discovery engine with potential for deep biological insights.

*(continue to page 163)*

*(continued from page 162)*

Here we describe the use of these screening approaches to identify genetic elements that regulate T cell effector differentiation, exhaustion, and stemness. Thousands of single and combined genetic perturbations were assessed using pooled in vitro screens. Promising candidates were evaluated using our automated, arrayed platform to characterize T cell expansion, tumor killing capacity, and the transcriptional cell state. Pooled and arrayed in vivo screens confirmed a set of novel perturbations that enhanced tumor control, potency, and persistence. Our high-throughput screening-based approach to cell design and optimization provides a platform that dramatically accelerates the development of safe and efficacious T cell therapies.

# cytoSNuBAR: A simple oligonucleotide-based method for multiplexing single-cell transcriptomics

Technology Development

POSTER 510

**Rui Ye[1,2], Kaile Wang[3], Yawei Qiao[2], Zhenna Xiao[3], Yun Yan[1,3], Jianzhong He[2], Nan Li[2], Yifan Wang[2], Min Hu[3], Shanshan Bai[3], Emi Sei[3], Steven H. Lin[1,2], Nicholas Navin[1,3,4]**

1 University of Texas MD Anderson Cancer Center, Graduate School of Biomedical Sciences, Houston, TX
2 University of Texas MD Anderson Cancer Center, Department of Radiation Oncology, Houston, TX
3 University of Texas MD Anderson Cancer Center, Department of Genetics, Houston, TX
4 University of Texas MD Anderson Cancer Center, Department of Bioinformatics and Computational Biology, Houston, TX

With the rapid development of high-throughput single cell RNA sequencing technologies, despite of the availability of several existing methods for sample multiplexing (eg: CITE-seq, MULTI-seq, ClickTag, etc. ), a flexible, highly scalable, and easy-to-use method remains greatly needed to reduce both batch effects and experimental costs. Here, we developed a cytoplasmic RNA simple nucleus barcoding (cytoSNuBAR) method that requires minimal starting cell materials and is readily adaptable to microdroplet platforms (eg. 10X Genomics). cytoSNuBAR utilizes unmodified oligonucleotide adaptors as sample barcodes by adding one additional tagmentation reaction step to the standard 10x cell library preparation protocol. We show that cytoSNuBAR can achieve highly efficient cell barcoding and high data quality that is comparable to standard 10x 3' chemistry data using both human breast cancer cell lines and normal human breast tissues. Additionally, we show that cytoSNuBAR is compatible with fixed cells, allowing great flexibility in sample storage and processing logistics. We also applied cytoSNuBAR to multiplex 32 mouse samples from syngeneic lung cancer models treated with 8 different combination therapies to investigate the molecular mechanisms of the sensitization effects of PARPi on chemoradiation (CRT) and immunotherapy. Our data show that PARPi combined with CRT or in combination with radiation and immunotherapy reprogrammed both the tumor cells and components of the TME, resulting in a more tumor-reactive phenotype. Our data suggest that cytoSNuBAR is a simple, flexible, highly-scalable and cost-effective sample-multiplexing method that achieves high cell barcoding efficiency and data quality in cell lines and tissue samples.

# Developing a personalized sequencing assay for cerebrospinal fluid liquid biopsies to detect minimal residual disease in children with high-risk central nervous system tumors

Technology Development

POSTER 511

**Sarah A Wilson[1], Adithe C Rivaldi[1], Olivia E Grischow[1], Jocelyn M Lucyshyn[1]. Lakshmi Prakruthi Rao Venkata[1], Kyle J Voytovich[1], James Fitch[1], Richard K Wilson[1,2], Elaine R Mardis[1,2,3], Margot A Lazow[2,4], Anthony R Miller[1], Katherine E Miller[1]**

1 The Steve and Cindy Rasmussen Institute for Genomic Medicine, Abigail Wexner Research Institute at Nationwide Children's Hospital, Columbus, Ohio, USA
2 Department of Pediatrics, The Ohio State University College of Medicine, Columbus, Ohio, USA
3 Department of Neurosurgery, The Ohio State University College of Medicine, Columbus, Ohio, USA
4 Pediatric Neuro-Oncology Program, Nationwide Children's Hospital, Columbus, Ohio, USA

Pediatric patients with high-grade central nervous system (CNS) tumors often experience disease progression, recurrence, or leptomeningeal dissemination despite treatment. Conventional techniques for disease surveillance (e.g., imaging and cerebrospinal fluid (CSF) cytology) lack sensitivity to detect low levels of minimal residual disease (MRD), which can inform prognosis and clinical decision-making. There is a critical need for robust, biomarker-driven assays to complement conventional approaches for disease evaluation.

Cell-free DNA (cfDNA) fragments are released during cellular turnover into biofluids including blood, saliva, urine, and CSF. Data on the utility of CSF liquid biopsies within neuro-oncology is limited and historically restricted to molecular techniques not translatable to pediatrics for two main reasons: Pediatric CNS tumors typically have individualized mutational profiles, and DNA yields from CSF are low, introducing a technical barrier. Therefore, we developed a personalized sequencing assay to detect tumor-derived CSF cfDNA as a surrogate for measuring MRD.

Preliminary experiments included isolation of cfDNA from small volumes (⩽ 2mL) of CSF and detection of mutations by high-depth (>50,000x) next-generation sequencing (NGS) at variant allele fractions (VAF) as low as 0.1%. Briefly, we used a targeted amplicon sequencing approach by designing primers flanking a tumor mutation (previously identified in the patient's exome sequencing) and then prepared amplicons for NGS using NEBNext® Ultra™ II DNA library preparation kit. In total, we profiled seven CSF samples from five unique patients and detected tumor-derived DNA in four, with VAFs ranging from 0.1%-4.5%. Amplicon sequencing demonstrated greater sensitivity than CSF cytological assessment, which was negative for each case.

*(continue to page 166)*

*(continued from page 165)*

Our current assay targets only one tumor variant for sequencing, thus limiting sensitivity. Although pediatric CNS tumorigenesis is sometimes driven by a single 'driver' oncogenic mutation, there are other 'passenger' mutations detected in exome sequencing which, although less relevant for understanding the diagnosis/prognosis, are still specific to the tumor, and therefore represent valuable targets for CSF profiling. We aim to develop a targeted hybridization capture assay to enrich specific DNA loci and enable detection of all tumor variants per patient, with the long-term goal of creating an actionable assay to assess treatment response and predict tumor progression, recurrence, and/or dissemination.

# Development of a fully automated high-throughput process for long-read sequencing library preparation



**Evan McDaid, Michelle Cipicchio, Ally Day, Corey Nolet, Roberto Luis-Fuentes, John Walsh, Catie Ramnarine, Erin LaRoche, Michael DaSilva, Katie Larkin, Jon Thompson, Mariela Mihaleva, Atanas Mihalev, Tom Howd, Niall Lennon, Stacey Gabriel**

Broad Institute of MIT and Harvard, Genomics Platform, Cambridge, Massachusetts, United States

Long read sequencing enables a deeper understanding of the most complex and challenging regions in the human genome, but until recently these library preparation methods have not been suitable for high throughput scale. Long reads are the gold standard for capturing structural variation, and are essential to building a clearer picture of the genetic basis for common disease.

We built a fully automated high throughput workflow that can produce Pacific BioSciences' HiFi data for hundreds of samples per month. This allows unprecedented population-scale analysis of structural variation in human whole genomes. The end to end workflow presented here addresses the challenges of handling high molecular weight DNA, complex and sensitive sequencing preparation methods, automated sample indexing and tracking, data review, and delivery.

Sample preparation is done entirely on a custom automation configuration using Hamilton Robotics Starlet instrumentation facilitating gentle sample handling. To streamline the process and limit the potential for human error, intervention by the user is limited.  Sample tracking is fully automated throughout every step of the process using pre-barcoded tubes and plates. This method enables record keeping of key sample metrics, reagent lot numbers, expiration dates, and other associated metadata that could impact sample outcome.

*(continue to page 168)*

*(continued from page 167)*

The inclusion of a fully automated process has allowed us to increase our scale 3.5 fold, from 39 cells a week to 140, while ensuring high quality results. We show that switching to an automated process does not hurt metrics in terms of turnaround time, reproducibility, and data quality. Additionally, we gained the ability to easily track and process multiplexed samples at scale reducing sequencing costs.

This automated workflow will allow researchers across the globe access to population scale long read studies, such as the All of Us Research Program. Continuous chain of custody, reagent tracking, and advances in the analytical pipeline have enabled a pathway for large scale clinical long read genomes and the potential to unlock previously unknown causes of human disease.

# Evaluating Alzheimer's disease–associated gene markers with Xenium In Situ Gene Expression

Technology Development

POSTER 513

**Michelli Faria De Oliveira [1*], Luigi Alvarado1, Sarah Taylor1**

1 10x Genomics, Pleasanton, CA, USA

*corresponding author

Understanding cellular and molecular states underlying neurodegenerative diseases is critical for new therapeutic strategies. The Alzheimer's disease (AD) field has made significant progress via next-generation sequencing technologies, such as single cell RNA-seq. However, the lack of spatial context may limit biological interpretations of AD, which is strongly linked to amyloid-beta (Aβ) deposition in the brain. Here, we evaluated AD-associated gene expression in the TgCRND8 mouse model and its wild-type (WT) counterpart, using the Xenium In Situ platform. A Xenium mouse brain panel including 248 genes covering major brain cell types was designed. We then designed a 99-gene add-on panel to the brain panel to include additional cellular state markers (e.g. activated and disease-responsive microglia, reactive astrocytes), plaque-associated genes, and AD-risk genes, curated from the literature. We selected three age-matched timepoints for WT and TgCRND8 mice, corresponding to sparse, moderate, and severe Aβ plaque burden. We placed coronal FFPE brain sections (5 µm) on 12 x 24 mm Xenium slides for deparaffinization and de-crosslinking. Next, DNA probes were hybridized to mRNA targets (347 genes). The DNA probes have two target binding regions and a gene specific barcode that identifies a target transcript. The target binding regions of the probe bind and ligate to form a circular DNA probe, which is then amplified by rolling circle amplification to generate many copies of the circularized probe. On the instrument, fluorescently labeled detection probes were hybridized to rolling circle products for a total of 15 cycles of in situ imaging to generate a unique optical signature for each gene. An on-instrument analysis pipeline decodes the fluorescent signals, which can be visualized with the Xenium Explorer software. We accurately mapped canonical neuronal layer markers previously described in the literature and annotated major brain cell types. By overlaying Xenium data with an Aβ immunofluorescence-stained image of a serial adjacent section, we mapped plaque-associated genes in the vicinity of Aβ plaques in TgCRND8, which were downregulated in the WT mice. Our results demonstrate that Xenium is a powerful tool to resolve AD-associated gene expression at both single cell and spatial resolution.

# High-throughput spatial-omics sample processing of archival FFPE specimens at the whole-transcriptome level using digital spatial profiling

Technology Development

POSTER 514

**Margaret Hoang,** **Katrina van Raay, Gokhan Demirkan, Joseph M Beechem**

NanoString® Technologies Inc, Research & Development, Seattle, WA, USA

Spatially resolved whole transcriptomes offer promising new insights into the study of human disease from archival FFPE specimens, including mechanism of drug action, biomarker discovery, and immune response. However, an emerging unmet need in spatial-omics is the ability to standardize and process patient samples from large cross-sectional or longitudinal cohort studies. Here we test batch slide processing and storage of whole transcriptome stained slides to maximize the run rate through one GeoMx® Digital Spatial Profiler (DSP) instrument. We processed 28 slides using the standard automated slide preparation protocol for in situ hybridization of 18,000+ probes to the whole transcriptome. For biosamples, we use standard reference control samples for the testing, comprised of 14 FFPE serial sections of a tissue microarray (TMA) with 10 different human tissue types and 14 FFPE serial sections of a cell pellet array (CPA) with 11 different human cell lines. After batch slide preparation, we ran the GeoMx DSP instrument with two slide replicates, each of the TMA and CPA, as Day 1 of the standard workflow with no storage and then stored the whole transcriptome hybridized tissues for subsequent runs on Day 2, 3, 4, 7, 14, and 21 as the test workflow with storage. Results show high concordance of gene expression counts of the 18,000+ probes between spatially matched regions-of-interest between all days with and without storage (Pearson R > 0.9). This concordance was similar to the observed slide-to-slide variation between replicates within each day. We also find throughout the storage time course a similar number of genes detected in tissues and similar fluorescence intensity of morphology marker antibody staining (PanCK, CD68, CD45). Furthermore, we show that gene expression profiles drive unsupervised hierarchical clustering of tissues and cell lines rather than days spent in storage. We observed a modest reduction of fluorescent intensity of the nuclear stain Syto13 with slide storage up to 21 days, but the nuclear stain remained functional with similar numbers outputted from the nuclei counting algorithm over time. Our results show a high throughput, scalable workflow that enables spatially resolved whole transcriptomes of large patient cohorts (100+) with archival FFPE tissue within less than a month.

# HyDrop 2.0: A scalable droplet microfluidics-based workflow for single-cell multiomics profiling

Technology Development

POSTER 516

**Poovathingal Suresh[1], Florian V De Rop[1,2], Gert J Hulselmans[1,2], Carmen Bravo González-Blas[1,2], Stein Aerts[1,2]**

1 VIB-KU Leuven/VIB Center for Brain & Disease Research, Belgium
2 Laboratory of Computational Biology, Department of Human Genetics, KU Leuven, Belgium

Single-cell RNA-sequencing and single-cell assay for transposase-accessible chromatin (ATAC-seq) technologies are currently used extensively to create cell type atlases for a wide range of organisms, tissues, and disease processes. Commercialized emulsion solutions (such as 10X Genomics) offer robust and reproducible workflow for profiling of thousands of cells. However commercial platforms offer limited flexibility for assay modifications and the cost aspects are prohibitive. To increase the scale of single cell sequencing, lower the costs and to pave way for more specialized single cell and multiome assays, custom droplet microfluidics offers attractive options.

We have developed HyDrop, a flexible open-source droplet platform to perform single cell indexing of both RNA and ATAC sequencing. Single cell indexing in HyDrop is achieved using dissolvable hydrogel beads synthesized using acrylamide polymerization chemistry. In contrast to other droplet workflows, we have observed that the dissolution of beads in reducing environment improves the barcoded primer release and enhances the overall reaction efficiency and yield. In earlier version of the assay, we developed DNA polymerase based three-stage extension for clonal barcoding of the hydrogel beads. However, we have observed that the polymerase extension method introduces higher barcoding error rates and in turn affects the sensitivity of the assay by lowering the fraction of reads in cell (FRIC). To overcome this, we have modified the barcoding strategy to a ligation-based workflow which has resulted in >50% improvement of the barcoding efficiency (FRIC score). For both RNA and ATAC seq workflows, we have performed a number optimization steps to find the most optimal performing combination of buffers, enzymes and additives.

*(continue to page 172)*

*(continued from page 171)*

To benchmark and compare the performance of HyDrop bead barcoding chemistries, we generated single-nuclei libraries from snap-frozen, dissected adult mouse brain cortex. Here we will present the results of this benchmarking done on both the RNA-seq and ATACseq assays. During the development of HyDrop, we have made several adaptations and modifications to the existing single cell/droplet workflows and the combination of these adaptations resulted in a significantly increased sensitivity compared to InDrop and Drop-seq methods, and making HyDrop more user-friendly and with a running cost of ~0.02 $ per cell.

# Identifying genetic variation in cancer using targeted long-read sequencing



**Shruti V Iyer[1,2], Melissa Kramer[1], Amber N Habowski[1], David A Tuveson[1], Sara Goodwin[1], W Richard McCombie[1]**

1 Cold Spring Harbor Laboratory, Cold Spring Harbor, NY;

2 Program in Genetics, Stony Brook University, Stony Brook, NY

Structural variations (SV) tend to be recurrent and have been associated with several cancer types. Next-generation sequencing (NGS) is mostly blind to large SVs, lacking sensitivity with false positive rates up to 89% in SV detection. Long-read sequencing generates read lengths of tens of thousands of bases and has helped identify thousands of genomic features pertinent to cancer previously missed by NGS. However, whole genome long-read sequencing sometimes does not generate adequate coverage of some alleles in heterogeneous samples like tumors.

Targeted sequencing improves accuracy and coverage by providing the depth necessary to detect rare alleles in a heterogenous population of cells. Recently, we developed ACME, an Affinity-based Cas9-Mediated Enrichment method, which is an improvement on nanopore Cas9-targeted sequencing (nCATS). Like nCATS, ACME is an amplification-free approach that can simultaneously assess genomic variations and DNA methylation from the same sequencing run. Importantly, ACME helps reduce background reads, increasing on-target coverage and size of target regions capturable. Using ACME to target 10 cancer genes in MCF 10A and SK-BR-3 breast cell lines, we saw an increase in target coverage by 2- to 25-fold compared to nCATS. ACME increased the number of end-to-end reads spanning 100% of gene targets by 3- to 7-fold, giving >60-fold target enrichment, 35-65x coverage, and 3-20 end-to-end reads spanning BRCA2, a 95kb target on our panel. Across our panel, we observed an increase in enrichment, depth, and number of end-to-end reads, which ultimately helps with better alignment and SV detection. ACME detected all SVs within our target regions previously inferred by PacBio and ONT whole genome long-read sequencing, but with higher depth, allowing for rare variant detection, making it an effective long-read targeting platform.

*(continue to page 174)*

*(continued from page 173)*

We are currently using ACME to uncover large variants of clinical significance in pancreatic ductal adenocarcinoma that may have been missed by NGS. Pancreatic cancer patient–derived organoids represent the full spectrum of disease, making them invaluable models to assess personalized treatment options. Using ACME to evaluate the mutation status of genes of interest in these organoids could help identify missing variants, enabling a better evaluation of predictive biomarkers.

# Implementation of BIOSCAN at the Wellcome Sanger Institute: A high-throughput biodiversity monitoring pipeline

Technology Development

POSTER 518

**Scott Thurston[1], Naomi Park[1], Lyndall Pereira da Conceicoa[1], Emma Betteridge[1], Andrew Sparkes[1], Abdulrahman Tuameh[1], Rhona Long[1], Mara Lawniczak[1], Ian Johnston[1].**

At the Wellcome Sanger Institute (Sanger), we have been developing a robust and streamlined automated pipeline to support a new, long-term biodiversity monitoring project. Sanger expects to receive up to 20,000 samples per month in the initial stages of this 5-year project, but the pipeline is readily scalable far beyond these numbers.

Insects are collected at collaborator sites using Malaise traps. Insect carcasses are then arrayed into 96-well plates for delivery to Sanger.

Sanger has a requirement to maintain insect carcass morphology for taxonomic identification after nucleic acid extraction, which poses significant operational complexities. In order to address this issue, Sanger has implemented a novel, automated non-destructive extraction process which aspirates lysates from around insect carcasses, minimising morphological damage. Positive and negative controls are also introduced into randomised lysate well locations during this automated process in order to identify and mitigate systemic issues, cross-contamination, and plate swap events.

Lysates proceed directly into a targeted amplicon pipeline; Sanger is able to omit purification steps for a streamlined approach by utilising polymerases that are highly tolerant of inhibitors and nanoliter-range liquid handling technology, which also minimises costs and maximises throughput.

The amplicon panel utilised by Sanger contains the highly conserved mitochondrial gene cytochrome c oxidase I as a bio-identification system, with additional targets in development for plants and fungi to capture diet, pollination and diseases of sampled organisms.

*(continue to page 176)*

*(continued from page 175)*

96x 96-well input plates are ultimately combined into a single pool of 9,216 samples utilising a combinatorial index approach. Unique combinations of 96 'forward' and 96 'reverse' indexed primers, are arrayed into sets of 24x 384-well plates using the Echo and Dragonfly platforms.

Sample pools are sequenced on the PacBio Sequel II, analysis and species identification within the highly-multiplexed pools is performed using the rapid and highly configurable mBRAVE software platform.

BIOSCAN - International Barcode of Life (ibol.org)
BIOSCAN – Wellcome Sanger Institute

# Implementation of high-throughput saturation genome editing at the Wellcome Sanger Institute



**Nicholas Redshaw[1], Michael Quail[1]**

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

Saturation Genome Editing (SGE) is a form of deep mutational scanning that uses CRISPR/Cas9-based technology to systematically alter each position in a target region to explore its function. SGE has the potential to revolutionise diagnostics and treatment in rare disease and cancer by providing large-scale functional classifications for variants of uncertain significance, including those yet to be observed in the clinic. Here we present work by the Wellcome Sanger Institute to develop a high-throughput SGE pipeline that will enable 100s-1000s of SGE screens to be performed per year with an ultimate aspiration to generate variant effect maps for all 20,000 genes in the human genome within a 10-year period as part of the Atlas of Variant Effects alliance.

# Joint genetic and epigenetic sequencing technology leads to improved genetics compared to existing methylation calling methods

Technology Development

POSTER 520

**Casper K Lumby**[1,2], **Páidí Creed**[1], **Eric Banks**[2], **James Emery**[2], **Michael Gatzen**[2], **Christopher Kachulis**[2], **Megan Shand**[2], **Nicholas Harding**[1], **Jamie Scotcher**[1], **Joanna D Holbrook**[1]

1 Cambridge Epigenetix Ltd, Cambridge, United Kingdom
2 Broad Institute of MIT and Harvard, Cambridge, MA, USA

There is more to DNA than just the genetic bases A, C, G and T. Epigenetics is of growing scientific interest and plays a causal role in cell fate, ageing, response to environment and disease development. The ability to accurately determine both genetic and epigenetic letters, and their interaction with each other, is important for deriving new biological insights.

In humans, methylation of cytosines represents the most common form of epigenetic letters. Current approaches for measuring methylation utilise pre-sequencing protocols for the conversion of unmodified cytosines to uracil. Upon sequencing, methylated cytosines may then be read off as C's, whilst unmodified cytosines are identified as T's. This approach generates valuable epigenetic insight, but sacrifices genetic determination in the process, as unmodified C's become indistinguishable from T's. The reduced genetic representation results in suboptimal read mapping and limited variant discovery.

CEGX 5-Letter seq enables the simultaneous measurement of genetics and epigenetics by expanding the number of information states in next-generation sequencing from four to sixteen. Of these sixteen states, five are reserved for measuring genetics (A,C,G,T) and methylation (modified cytosine), whilst the remaining eleven are utilised for error-suppression. This approach allows for direct, digital and phased discrimination of genetic and epigenetic letters on the same read.

*(continue to page 178)*

*(continued from page 178)*

From the evaluation of Genome in a Bottle samples using multiple methylation sequencing methods, we show an increased genetic accuracy in 5-Letter seq compared to established technologies. The improved accuracy results from enhanced error suppression and the ability to directly measure all four genetic bases. Importantly, the capacity to discriminate C and T bases provides a significant advantage over competing technologies.

Additionally, we demonstrate how accuracy improvements in 5-Letter seq translate into enhanced calling of germline single nucleotide variants. The inherent error-correction results in increased precision, whilst the ability to capture complete genetics leads to a consistently high sensitivity across all twelve substitution types (A>C, A>G, …,T>G), including C>T, which is the most common mutation in the mammalian genome. Overall, 5-Letter seq provides a previously unseen level of genetic accuracy in a methylation calling technology.

# MAGNA: A single-molecule analysis platform for the capture and interpretation of complex information from the 'dynamic genome'

Technology Development

POSTER 521

**Jimmy Ouellet[1], Thibault Vieille[1], Francois Paillier[1], Nigel Skinner[1] and Gordon Hamilton[1]**

1 Depixus SAS, 3-5 Impasse Reille, 75014 Paris, France

Despite the enormous technological advances made in genomics over recent years, the latest methods and tools still fall short in their ability to capture critical epigenetic and structural data from DNA and RNA. Depixus' MAGNA™ technology will enable research to move from a world of measuring nucleic acids as linear, one-dimensional sequences, to one of dynamic, three-dimensional objects with many base modifications, and at scale.

MAGNA™ is a novel single molecule genomic analysis platform based upon magnetic tweezer technology. Large numbers of individual nucleic acid molecules can be captured in their native form, immobilised within a flow cell, and are then available for repeated, non-destructive interrogations that can reveal a broad range of features including base modifications, molecular structure, and interactions with other nucleic acids, proteins, or drugs. These capabilities unlock previously unattainable insights into the "dynamic genome" – the layers of information beyond the four-base coding sequences of DNA and RNA, that are key to gene expression, regulation, and control.

Furthermore, with MAGNA™ it is possible to measure both the on- and off-rates of ligands that bind to DNA or RNA and to identify where such binding interactions occur. These capabilities can be used to analyse the binding characteristics of drugs or proteins that target nucleic acids.

We present data from recent work that shows how we can directly characterize the epigenetic landscape of specific genes, enriched from human genomic DNA, while avoiding the need for sample amplification or chemical transformation. We will also discuss the broader application of MAGNA™ in translational research and drug discovery.

# Molecular spikes: Ground truth molecule counting for single-cell RNA-seq

Technology Development

POSTER 522

**Christoph Ziegenhain[1] *, Gert-Jan Hendriks[1]*, Michael Hagemann-Jensen[1], Rickard Sandberg[1]**

1 Department of Cell & Molecular Biology, Karolinska Institutet, Sweden
*Equal Contribution

Counting of RNA molecules using unique molecular identifiers (UMIs) is ubiquitous in single-cell RNA sequencing (scRNA-seq) in order to account for amplification biases. We present molecular spikes [1], the first tool to experimentally evaluate molecule counting performance. By engineering RNA spikes of various properties (length, GC content) with inbuilt UMIs, we identify critical experimental and computational conditions for accurate RNA counting in scRNA-seq experiments. We uncover impaired RNA counting in methods where UMI counts were inflated and not informative of molecular abundances. Furthermore, we leverage molecular spikes to improve estimates of total endogenous RNA amounts in cells, and introduce a strategy to correct experiments with impaired RNA counting. We show that molecular spikes can also be used to optimize computational methods for the analysis and error-correction of sequencing data containing UMIs. The molecular spikes are made available for the research community (https://github.com/sandberg-lab/molecularSpikes) along with the accompanying R package UMIcountR (https://github.com/cziegenhain/UMIcountR). We anticipate that wide-spread implementation of molecular spikes will improve the validation of new methods and help to better estimate and adjust for cellular mRNA amounts as well as enabling more in-depth characterization of RNA counting in scRNA-seq.

[1] Ziegenhain et al., Nature Methods, 2022

# Multi-sample multi-omics preparation on a novel digital micro-fluidic system

Technology Development

POSTER 523

**Presenting Author Candice Wike[1] and Ritin Sharma[2,3], Dean Ellis[1], Melissa Martinez[2,3], Cassie Schumacher[4], Stephanie Pond[1] and Patrick Pirrotte[2,3]**

1 Translational Genomics Research Institute (TGen), Emerging Technologies, Phoenix, AZ, USA
2 Translational Genomics Research Institute (TGen), Cancer and Cell Biology Division, Phoenix, AZ, USA
3 Translational Genomics Research Institute (TGen), Integrated Mass Spectrometry Shared Resource, Phoenix, AZ, USA
4 Miroculus, Inc., San Francisco, CA USA

Characterization of the proteome is key to connecting genotype to phenotype and is a powerful tool for the identification and validation of novel biomarkers. The Collaborative Center for Translational Mass Spectrometry is a leading research facility specializing in ultra-deep proteomics profiling and enables data integration across proteomics, metabolomics, genomics and transcriptomics to provide a united and actionable overview of tumor biology in precision medicine. Sample preparation for mass spectrometry today is labor intensive, with limited options for scaling due to the organic chemicals used and adsorption of proteins to many surfaces, particularly at low concentrations. There is a need for a system that allows for robust and consistent generation of high-quality protein digests for use in mass spectrometry with minimal manual effort. Here we describe our custom sample preparation method for mass spectrometry on the digital micro fluidic (DMF)-based system, Miro-Canvas (Miroculus). We have tested processing up to 12 samples in parallel while achieving accurate, reliable and reproducible digests with minimal cross-sample contamination.

# Multiplexed phasing of clinically relevant long human PCR amplicons with short reads



**Jessica Smith**
1 seqWell, Inc

While short-read sequencing technology can call variants with high fidelity, phasing variants over gene-length regions remains a challenge. Long-read sequencing methods are better suited for this application, but have inadequate accuracy and are difficult to scale to many samples. We have developed a novel library preparation technology that links short reads across long DNA fragments and multiplexes many samples for fast and affordable parallel phasing of dozens of gene-length amplicons.

We used six kilobase amplicons of the gene Cyp21A2 as a demonstration of our technology because for best diagnostic accuracy, genotyping of Cyp21A2 must fully resolve the variants on both alleles. Variants in this gene are associated with congenital adrenal hyperplasia caused by 21-hydroxylase deficiency. Pathogenic variants in Cyp21A2 correspond to disease states ranging from mild to severe and affected individuals often have two different pathogenic alleles. To be useful as a clinical or prenatal diagnostic, genotyping of Cyp21A2 would benefit from a haplotype assembly solution that is affordable and scalable, and amenable to widely available short-read sequencing platforms.

We generated amplicons from individuals with known haplotypes of the Cyp21A2 gene. Our workflow tags individual samples with sample-specific barcodes prior to pooling, then adds a phasing barcode at multiple locations along long DNA fragments. We multiplexed 24 samples generated with four different input mass levels across an Illumina MiSeq flowcell, resulting in roughly 737,000 reads per sample. Phased data from these samples showed an average of 2.9 linked reads per phasing barcode with average fragment insert size of ~220 base pairs. After calling variants on aligned sequencing reads using standard bioinformatics tools, heterozygous variants were linked via the phasing barcode and haplotypes assembled using maximum likelihood estimation. Heterozygous variants at 20 positions were phased correctly at all positions for > 95% of samples tested.

The pooled format of the phase barcoding process presents significant workflow advantages for processing large numbers of samples with an efficient and scalable workflow. We anticipate this technology will have significant potential to be extended to other applications including single cell sequencing, bacterial and viral metagenomics, and full length transcript assembly.

# mutREAD is a novel, cost-effective, FFPE-compatible method to study the evolutionary dynamics of tumors



**Karol Nowicki-Osuch1, Aisha Yusuf2, Wladyslaw Januszewicz3, Rebecca C. Fitzgerald2**

1 Columbia University, Irving Institute for Cancer Dynamics, New York City, NY, United States
2 Cambridge University, Early Cancer Institute, Cambridge, United Kingdom
3 The Centre of Postgraduate Medical Education, Warsaw, Poland

Two major mutational processes shape the evolution of cancer. Cells either acquire individual point mutations (single cell nucleotide variants – SNVs) or large genomic regions undergo gains or losses of DNA (copy number alterations – CNAs). Recent discoveries indicate that individual mutagens leave specific footprints (mutational signatures) that are encoded in the tri-nucleotide context of SNVs. Less is known about the processes driving patterns of CNAs. Whole Genome Sequencing (WGS), albeit expensive, is the gold standard method used to detect mutational signatures, SNVs, and CNAs. Recently, we demonstrated that mutREAD (mutational signature detection by restriction enzyme-associated DNA sequencing) is a cost-effective WGS alternative that accurately identifies SNVs and mutational signatures associated with gastrointestinal cancers.

Here, we developed a new computational pipeline that comprises of: 1) a new statistical model to identify relative genomic coverage within mutREAD data, 2) a DNA segmentation framework that identifies relative CNAs in mutREAD data, 3) a single nucleotide polymorphism (SNP)-based approach for absolute copy number calculation, and 4) a statistical framework for identification of clonal and subclonal SNVs in the mutREAD data.

When applied to a cohort of 20 esophageal adenocarcinoma cases where WGS and mutREAD data are available, we observed very good concordance in the CNAs calls between the methods. Further extension of the method to multi-regional, micro-dissected samples (6 cases), collected from esophageal squamous cell carcinoma samples preserved in FFPE material, showed that mutREAD analysis can accurately reconstruct the evolutionary trajectory of individual tumors. Indeed, CNA and SNV-based phylogenic links were identified between tumor and pre-cancerous lesions confirming their natural history.

*(continue to page 185)*

*(continued from page 184)*

Our analysis demonstrates that mutREAD allows for the study of mutational signatures, CNA analysis, and reconstruction of tumor evolutionary dynamics in large patient cohorts. Furthermore, the method is compatible with formalin-fixed paraffin-embedded samples, opening a new avenue for studies of clinical-relevant samples in a cost-effective manner.

# Nanopore-based sequencing characterizes the complex rearrangement and methylation landscape of KMT2A gene fusions in acute myelogenous leukemia

Technology Development

POSTER 526

**GiWon Shin[1], Bo Zhou[2], Raegan N. Wood[3], Hanlee P. Ji[1]**

1 Division of Oncology, Stanford University School of Medicine, Stanford, CA, United States, 94305
2 Department of Psychiatry and Behavioral Sciences, Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305
3 Biological Sciences Department, Dixie State University, St George, UT 84770

In acute leukemia, chromosomal translocations are important drivers of cancer and have prognostic significance. KMT2A translocations and the resulting gene fusions are among the most common chromosomal abnormalities in acute leukemias. However, very little is known about their rearrangement structure because of different translocation sites. In addition, nothing is known about how these translocations alter the methylation context of the affected chromosomal segments.

We use a high molecular weight (HMW) DNA enrichment strategy combined with nanopore sequencing to characterize these translocations. This technology, termed as CRISPR/Cas9-targeted ultra-long read sequencing (CTLR-Seq), was developed using customizable Cas9 nucleases to isolate large genomic DNA targets from intact cells for long read sequencing. This approach enriches and extract target regions that are as large as 1 Mb that extend over a translocation. The HWM DNA target undergo nanopore sequencing to determine the rearrangement structure and extract the overlapping methylation landscape.

*(continue ro page 187)*

*(continued from page 186)*

We used this approach to characterize KMT2A translocations from a set of acute myelogenous leukemias. These samples were previously tested with clinical cytogenetic methods. We discovered gene fusions between KMT2A and five gene partners on other chromosomes. This analysis identified the breakpoints at base pair resolution, generated haplotypes with end-to-end contig assembly of wildtype and fusion alleles, and revealed the leukemia-specific epigenetic surrounding the translocation locus. With the methylation data we predicted the expression activity of the fusion and wild type genes. This information can be obtained only when wild type and fusion genes are separated, and the entire translocation locus is isolated as an intact HMW DNA. We present examples of epigenetic regulations on fusion and wild type genes with results from leukemia samples.

# Novel high-throughput single-cell analytical methods using Pre-templated Instant Partitions (PIPseq)

Technology Development

POSTER 527

**Adam Abate,** UCSF

Single cell transcriptomics has revolutionized modern biology by providing unprecedented insight into the mechanisms of cellular differentiation and function in complex tissues and organisms. As scRNA-seq transitions from a discovery tool into translational applications requiring complex experimental design, multiple sample comparisons, and multivariable sample perturbations, there is a growing need for simple, scalable, and cost-effective library preparation methods. Fluent BioSciences has developed new methods for scRNA-seq enabled by Pre-templated Instant Partitions (PIPseq), for simple, sensitive, and scalable scRNA-seq without the need for expensive instrumentation and microfluidic consumables. We demonstrate single-tube PIPseq configurations accommodating 200,000 - 1 million cells and evaluate key performance metrics using model cell lines, peripheral blood mononuclear cells, and complex tissues. PIPseq enables novel experimental paradigms not addressable with current platforms including identification of very rare cell populations from complex mixtures, surface epitope mapping, and multiplexed sample batching.

# NovaSeq X and NovaSeq X Plus Sequencing Systems are powered by Illumina's revolutionary XLEAP-SBS Chemistry enabling higher density sequencing and faster runtime with improved usability and sustainability

Technology Development

POSTER 528

**Johanna Whitacre**, **Ros Jackson, Michael Chesney, Alix Kwasniewska, Amanda Jackson, Suzi Bryan, Daniel Ortiz, Elliot Lawrence, Tim Davis, Vikki Evans, Patti Iavicoli, Steven Yun, Antoine Francais, Misha Golynskiy, Claire Bevis-Mott, Lea Pickering, Valerie Uzzell**

Illumina, Inc., Research and Development, San Diego, CA, USA

NovaSeqTM X and NovaSeq X Plus Sequencing Systems deliver major chemistry innovations designed to enable high quality sequencing at ultra-high flow cell densities with faster turnaround time, greater simplicity and significant improvements in sustainability. Illumina XLEAP-SBSTM chemistry introduces XLEAP-SBS nucleotides containing novel block and linker chemistry that is paired with our new XLEAP-SBS polymerase engineered for high fidelity and speed. Novel nucleoside block greatly improves stability by 50x-500x and enables faster deblock kinetics. The greatly improved nucleotide stability and much faster incorporation and deblock kinetics enable much faster SBS cycle times. Newly redesigned clustering reagent and recipe enable clustering at higher density flow cells with smaller feature sizes and allows on-board library loading of 8 individually addressable lanes with a 4-fold reduction in library input relative to NovaSeq 6000 automated onboard clustering. Additional NovaSeqTM X reagents were redesigned for improved stability, with the introduction of new lyophilized and thermostable reagents to allow kits to be shipped ambiently to customers, eliminating the need for dry ice or cold packs, greatly reducing waste and improving sustainability.

# Optimization of DNA library preparation workflows for emerging sequencing platforms



**Pingfang Liu, Chen Song, 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 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 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. Modular and tunable library preparation approaches will further facilitate workflow management and flexible reagent inventory for users working with multiple sequencing platforms.

# Accurate and simultaneous measurement of genetics and epigenetics with five-letter sequencing

**Páidí Creed[1], David Currie[1], Nick Harding[1], Casper K Lumby[1], Jack Monahan[1], David M Morley[1], Fabio Puddu[1], Jamie Scotcher[1], Rosie Telford[1], Jean Teyssandier [1], Shirong Yu[1], Joanna D Holbrook[1]**

1 Cambridge Epigenetix Ltd, Cambridge, United Kingdom

The interaction of genetics with the DNA epigenome is of growing scientific interest and plays a causal role in cell fate, ageing, response to environment and disease development. The ability to accurately determine both genetic and epigenetic letters, and their interaction with each other, is important for deriving new biological insights.

In humans, chemical modification of cytosines, such as 5-methylcytosine, represent the most common form of epigenetic letters and its combination with genetic changes has been shown to be more sensitive for disease detection than either genetics or epigenetics alone and to constitute the causal path for genetic variants associated with disease susceptibility.

Current approaches for determining modified cytosines apply next-generation sequencing in combination with a pre-sequencing protocol that differentially converts genetic bases dependent on epigenetic status. However, the ability to discriminate epigenetic state comes at a cost of incomplete genetic determination; inherently, it is impossible to determine full genetics and epigenetics from a system limited to only four states of information. To compensate multiple analyses can be performed on aliquots from the same sample. However, multiple analyses are not only expensive in terms of sample volume, time and reagent costs, but also yield suboptimal information due to the inherent inaccuracy of data integration and the loss of phasing between genetic and epigenetic letters.

*(continue to page 192)*

*(continued from page 191)*

CEGX 5-Letter seq addresses these challenges by expanding the number of information states in next-generation sequencing from four to sixteen, and then resolving these to obtain reads in 5-letter space (A, C, G, T and modified C). Here we present comprehensive evaluation on data derived from the application of CEGX 5-Letter seq to public reference materials, illustrating high quality genetic and epigenetic calls and robust performance across multiple distinct biological samples. Furthermore, we use the phased nature of the technology to study the prevalence of allele-specific methylation across all seven Genome in a Bottle samples and make this data freely available for further study.

# Protein–DNA interactions at the bench: CUT&RUN/Tag with nanopore sequencing

Technology Development

POSTER 531

**Paul W. Hook**[1] **and Winston Timp**[1]

1 Department of Biomedical Engineering, Johns Hopkins University

Exploring protein-DNA binding with sequencing is key to a greater understanding of gene regulation and nuclear organization. Two popular methods for profiling protein-DNA interactions include Cleavage Under Targets and Release Using Nuclease (CUT&RUN) and Cleavage Under Targets and Tagmentation (CUT&Tag). CUT&RUN uses micrococcal nuclease to cleave DNA around bound proteins, while CUT&Tag uses a transposase to insert adapters around bound proteins. Though typically they are used with short-read sequencing platforms, this has limitations in cost of equipment, limited read lengths, and time needed to complete sequencing runs. Motivated by these limitations, we tested CUT&RUN and CUT&Tag on the Oxford Nanopore Technologies (ONT) long-read minION platform due to its small physical footprint, lack of read length limitations, and the ability to perform real-time analysis.

We first performed CUT&Tag experiments on the GM12878 cell line. We targeted histone marks (H3K27me3 and H3K4me3), a transcription factor (CTCF), and a negative control (IgG), finding high concordance with existing short-read data. We then performed CUT&RUN without amplification on the same targets. Cut fragments diffusing out of the nucleus are purified during CUT&RUN and can be sequenced without amplification, allowing for protein-DNA binding and native CpG methylation detection with nanopore without additional chemical or enzymatic treatment. The sequenced fragments showed expected nucleosomal laddering including tetra- and pentanucleosomes (>600 base pairs), which cannot be sequenced on short-read platforms. We found H3K4me3 and H3K27me3 samples showed high concordance with short-read profiles, while CTCF profiles could not be initially recapitulated. Optimization of a variety of targets is on-going. For H3K4me3 and H3K27me3 samples, we called native CpG methylation, demonstrating CUT&RUN can be transformed into a real-time, multimodal, single-molecule assay without additional molecular biology techniques.

By combining CUT&RUN and CUT&Tag with nanopore sequencing, we have enabled affordable benchtop sequencing and measurement of protein-DNA interactions. This combination will expand the portability of protein-DNA binding assays, increasing accessibility to these assays in a range of environments.

# Rapid whole-slide spatial analysis of FFPE tissues with true multiomic panels enables the discovery of key cellular niches

Technology Development

POSTER 532

**Julia Kennedy-Darling**, **Akoya Biosciences**

Spatial biology analyses are critical to understanding key cellular interactions. While there have been tremendous advancements in spatial technologies, there was always a trade-off between plex, resolution, and speed. Akoya's PhenoCycler®-Fusion spatial biology platform solves this tradeoff by enabling rapid spatial phenotyping of whole FFPE slides for the detection of protein targets, RNA targets, or a combination of both from the same tissue. The workflow is inherently unbiased, enabling entire tissue resections to be analyzed in the detection moiety of choice, without compromising on data quality. Here we show the versatility of the technology in enabling the detection of more than 100 targets in FFPE tissues across a variety of human samples. The high-throughput capability of the platform, coupled with data compression makes analyzing whole slide sections routine and opens the door to cohort analysis for spatial phenotyping. By detecting 100+ targets and mapping every cell, new tissue structures and functions were discovered previously undetected by sequencing methods and traditional IHC readouts. The uniqueness of the PhenoCycler chemistry to detect both RNA and protein targets on a tissue section at the same time means critical questions related to signaling pathways, immune infiltration, and tumor environment can be answered directly using the optimal target panel. Multiomic spatial phenotyping is typically performed using a variety of cell differentiation markers and other key protein targets that accurately identify cells of interest. In contrast, questions of signaling pathways and secreted protein expression are addressed using spatial RNA detection. The ability to measure both types of targets in a scalable manner across entire tissue sections makes any spatial biology question accessible.

# Redesign and automation of NGS workflows that allows scalability, rapid turnaround, flexibility, and high efficiency

Technology Development

POSTER 533

**Bert van der Zwaag[1], Patrick van Zon[1], Koen van Gassen[1], Martin Elferink[2], Robert Ernst[2], Esther Janson[1], Carola Roelands[1], Hanneke van Deutekom[2], Marjan van Kempen[1], Hans Kristian Ploos van Amstel[1], 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.

In the past decade, the genome diagnostics section of the UMC Utrecht department of Genetics has seen annual growth of NGS based molecular diagnostic testing in excess of 10%, resulting in ~8.000 individual diagnostic WES samples processed in 2022. In 2023 a further increase in numbers of WES samples is foreseen with the transition of SNP-microarray based CNV detection and Sanger single gene resequencing to WES based analyses. In addition, the demand for rapid diagnostic testing (<10 days) in the pre- and perinatal setting has put strain on the systems implemented so far. To accommodate the demand for rapid analysis and volume increase and anticipating on projected additional growth in coming years, we adopted a modular workflow setup and implemented a series of optimizations and automations in our laboratory. We introduced liquid handlers with integrated focused-ultrasonicators (Tecan Fluent 480 with Covaris R230) to dilute, normalize and shear genomic DNA samples with minimal hands-on time. Furthermore, fully walk-away WES enrichment automation using the MagnisDx (CE-IVD, Agilent) provides the needed flexibility. Sequencing is performed on NovaSeq6000 systems (Illumina) and initial data analysis is performed on site using in-house developed bioinformatic workflows. Annotation and interpretation are performed using Alissa Interpret genomic data analysis software (Agilent).

*(continue to page 196)*

*(continued from page 195)*

This results in a streamlined setup enabling us to perform WES based diagnostics in >10.000 individuals annually (including rapid pre- and perinatal cases) with consolidation of high quality (>95% Exome coverage at >20X, >98% SNV and >93% Indel sensitivity, and robust CNV detection) and a significant reduction in dedicated technical support (-20% fte) and failed WES (<2%). Future projects will be aimed at further optimizing throughput and efficiency, reducing turn-around to facilitate ultra-rapid diagnostics (<5 days) and preparing our lab setup for routine diagnostic WGS analyses.

# scTaILoR-seq: Single-cell targeted isoform long-read sequencing

Technology Development

POSTER 534

**William Stephenson[1], Ashley Byrne[1], Dan Le[1], Kostianna Sereti[2], Karen Gascoigne[2], Tim Sterne-Weiler[3], Zora Modrusan[1]**

1 Genentech, Microchemistry Proteomics and Lipidomics + Next Generation Sequencing, South San Francisco, CA, USA
2 Genentech, Discovery Oncology, South San Francisco, CA, USA
3 Genentech, Bioinformatics, South San Francisco, CA, USA

Single-cell RNA sequencing methods predominantly employ short-read sequencing to characterize cell states and dynamics. While this approach is powerful for determining gene-level expression, full-length transcriptome complexity is still relatively unexplored at the single-cell level. Single-cell long-read sequencing enables the detection of RNA isoforms, but is hampered by lower throughput and requires mitigation of artifacts derived from template-switching based reverse transcription. To address these short-comings, we developed a two-stage hybridization-capture strategy called scTaILoR-seq (Single-cell Targeted Isoform Long-Read Sequencing). First, we target genes of interest using a commercially available gene enrichment panel. Second, we perform an enrichment step to mitigate artifactual molecules common to single-cell full-length cDNA preparations. We validate this approach by identifying and quantifying isoforms from mixed ovarian cell lines. We further expanded this approach utilizing distinct gene enrichment panels to characterize both immune and cancer related isoforms at the single-cell level from dissociated ovarian tumor samples. scTaILoR-seq establishes a blueprint for targeted single-cell isoform characterization and discovery in a manner that is tailored to long-read sequencing platforms.

# See everything: Flipping the paradigm on RNA sequencing through Level-Up full-length cDNA normalization

Technology Development

POSTER 535

**Richard I. Kuo**[1], **Yuanyuan Cheng**[1], **Björn Geigle**[1], **Juan Carlos Entizne**[1], **Bethany Nicholls**[1], **and Mark Barnett**[1]

1 Wobble Genomics, Edinburgh, Scotland, UK

The transcriptome presents as an ever changing sea of complexity that has eluded our previous attempts to explore its depths. One primary reason for this is the natural tendency for RNA libraries to contain a subset of relatively highly expressed transcripts which make it difficult to pick up other lower abundance transcripts. It is typically these lower abundance RNA that we wish to explore. Previously, the only two solutions to this issue was to either create probes to target low abundance transcripts or to simply sequence deeper. The former solution does not work for discovery purposes and the latter is cost prohibitive and in some cases does not provide ideal results.

With the advent of different types of RNA sequencing such long read RNA sequencing and single cell RNA sequencing, the objectives have now shifted toward being able to sequence with the greatest transcriptome coverage. This is in order to identify novel transcripts or produce a fuller picture of gene expression from individual cells. Given that sequencing is essentially a random sampling of a cDNA library, the best way to maximize the probability of sequencing all unique transcript in a library is for all unique transcript sequences to be represented at equal relative abundances.

We have developed a novel form of full length cDNA normalization (termed Level-Up normalization) that allows for the production of a cDNA library that maximizes the ability to sequence all unique transcripts in a library. Essentially Level-Up makes it feasible to sequence practically all unique transcripts in a sample.

In this presentation, we demonstrate the use of our Level-Up normalization technology for long read RNA sequencing with challenging sample types to improve transcriptome discovery and for single cell RNA sequencing to improve cell identification and rare cell discovery.

# Single-nuclei targeted long-read sequencing to explore splicing dynamics of _ScnXa_ gene family in mouse neurons



**Sheridan Cavalier, Siqi Chen[1], Stephanie Felker[2,3], Gregory M. Cooper[2], Richard M. Myers[2], Lindsay Rizzardi[2], Winston Timp[1,4]**

1 Johns Hopkins School of Medicine, Department of Molecular Biology and Genetics, Baltimore, MD, USA
2 HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA
3 University of Alabama in Huntsville, Department of Biotechnology, Huntsville, AL, USA
4 Johns Hopkins University, Department of Biomedical Engineering, Baltimore, MD, USA

Alternative splicing is highly dynamic in the brain, but accurately quantifying transient splicing events with short-read single cell transcriptomics is challenging. A key example is the inclusion of poison exons (PEs) which result in a premature termination codon triggering nonsense-mediated decay of the mRNA. The sodium voltage-gated alpha subunit gene family (ScnXa) includes several genes that have highly conserved PEs which serve as an important transcriptional regulation mechanism during embryonic mouse brain development. One human variant increasing 20N PE inclusion in SCN1A gene was characterized by Voskobiynyk et al. (2021) in a Scn1a(+/KI) mouse model with reduced Scn1a expression and seizure-related phenotypes. The 20N PE has implications in human disease, including Dravet syndrome, an early onset epileptic encephalopathy. Despite evidence suggesting cell-type specific regulation and expression of PEs throughout development, PE inclusion in the ScnXa family has only been examined via bulk short-read RNA-seq methods. These approaches fail to 1) distinguish transcript expression across distinct cell types in the brain; or 2) resolve exon-exon connectivity directly on the same cDNA molecules.

*(continue to page 200)*

*(continued from page 199)*

Single-molecule nanopore sequencing is well suited to describe the splicing dynamics and PE inclusion events of ScnXa genes as these long reads span entire transcripts allowing identification of cryptic splice sites, allele-specific expression patterns, and isoform diversity quantification. As ScnXa genes are long (110-180 kb) and lowly expressed specifically in neurons with complex spliced isoforms (1-13 kb), we performed targeted nanopore sequencing of 10X Genomics single-nucleus full-length cDNA libraries on sorted NeuN+ nuclei derived from the cortex of Scn1a(+/+)and Scn1a(+/KI) mice. We used a custom probe set to enrich for this gene family to 1) search for novel isoforms and PEs and 2) resolve neuronal sub-type differences in expression of ScnXa isoforms and PE inclusion. This approach enables us to detect rare splicing events in our single nuclei data and allows multiplexing of biological replicates on a single flow cell to improve confidence in novel PE and transcript identification. We anticipate this targeted single-nucleus long-read approach to be broadly applicable to other important gene families with complex splicing patterns over a wide range of expression levels.

# Single-cell profiling of the zebrafish brain and retina using combinatorial indexing

Technology Development

POSTER 537

**Frank Steemers,** **Patrick Boyd, Maggie Nakamoto, Nathan Pacheco, Lin Lin, Jerushah Thomas, Dmitry Pokholok, Felix Schlesinger, Toumy Guettouche**

Single-cell sequencing technologies have allowed for in-depth analyses of neuronal cell types and demonstrated the vast heterogeneity within these tissues. However, these studies are often limited in cell and sample throughput due to prohibitive costs. In this study we show that combinatorial indexing can be used to drastically increase cell output from a single experiment while simplifying sample multiplexing, significantly reducing bench time and cost per cell.

Briefly, we increased the throughput of on-market systems by adding upstream combinatorial indexing technology using the ScaleBio ATACseq pre-indexing kit. In this workflow nuclei from up to 24 samples are distributed across a plate for barcoded tagmentation, introducing an additional cell-based index to each molecule. Pooled nuclei can then be superloaded onto existing single-cell capture systems with up to 100,000 nuclei per channel, while multiplets can be resolved and recovered for downstream analysis using the tagmentation barcode for demultiplexing.
We demonstrate this technology first using a human/mouse barnyard experiment in which human and mouse cell lines were mixed prior to tagmentation, follow by loading of 100,000 nuclei into the capture system. Analysis of these data show effective recovery of nuclei with low effective doublet rates (<5%), clean distinction of human and mouse populations, and no decrease in sensitivity or increase in background with increasing nuclei per droplet.

Next nuclei from both retina and brain samples from 3 zebrafish were split across 24 wells on the ScaleBio tagmentation plate, using the tagmentation barcodes to tag individuals and tissues, and proceeding with the 100,000 cell load. Following standard QC filtering we show that all major cell types from both retina and brain can be identified, and the high cell numbers recovered here allow for identification of low-abundance or sample-specific cell types, uncovering new heterogeneity. Furthermore, profiling of all samples in a single experiment enabled direct comparison between individuals with no batch effect. These data show that combinatorial barcoding can be effectively applied to neuronal tissues, reducing bench time and cost without compromising data quality.

# Spatial analyses of the single immune cell surface proteome by molecular pixelation

Technology Development

POSTER 538

**Simon Fredriksson**[1,2], **Filip Karlsson**[1], **Alvaro Martinez Barrio**[1], **Tomasz Kallas**[1], **Maud Schweitzer**[1], **Emma Lundberg**[2,3], **Frida Björklund**[3], **Jose Fernandez Navarro**[1], **Louise Leijonancker**[1], **Divya Thiagarajan**[1], **Sylvain Geny**[1], **Erik Pettersson**[1], **Johan Dahlberg**[1], **Michele Simonetti**[1], **Prajatka Sathe**[1]

1 Pixelgen Technologies AB, Stockholm, Sweden
2 Royal Institute of Technology, CBH, Stockholm, Sweden
3 Stanford University, Stanford, CA, USA

The spatial arrangement of the surface proteome of single cells determines their activity in health and disease. Protein spatial architecture enables cell-cell communication, mobility, structure, and immunological activities such as the immune synapse. However, tools for studying these fundamental processes of life at the scale, resolution and throughput needed to facilitate large studies and data driven science are lacking, mainly due to the limitations of microscopy. To enable such studies, we have developed a DNA sequence based method to report the relative location of cell surface proteins by binding these with DNA-tagged antibodies and interrogating their relative locations using Molecular Pixelation. Molecular Pixelation uses DNA pixels (100 nm scale UMI concatemers), that hybridize to 5-10 cell bound target antibodies and imprint a UMI on these by a gap-fill ligation reaction effectively colocalizing target proteins. Subsequently, the relative location of these colocalization zones are determined by a second gap-fill ligation reaction using a second set of DNA pixels. The resulting product is amplified by PCR and subjected to standard short-read sequencing. All steps of Molecular Pixelation are performed in a reaction tube with no need for specialized equipment. Our pipeline computationally arranges the sequencing reads into spatial single cell proteome information of around 1000 pixels per cell, depending on cell type.

*(continue to page 203)*

*(continued from page 202)*

We have analyzed various immune cell types under a range of conditions and found novel patterns of surface proteome architecture by assigning proteins a Polarity Score, and protein combinations with Colocalization Scores, for each single cell. The Polarity Score reflects a protein's level of non-random distribution on a cell surface. It can be exemplified by the polarization of CD3 on T-cells upon activation leading to the initiation of the immune synapse. The immune synapse governs the resulting activity between T-cell and antigen presenting cell interactions.

Molecular Pixelation is a novel addition to our single cell analysis capability to reveal new insights into cellular life. The method is also under development for FFPE tissue samples to provide throughput, resolution and multiplexing for spatial tissue proteomics, beyond what is achievable with traditional imaging using microscopes and light.

# Spatially resolved single-cell transcriptomic imaging in formalin-fixed paraffin-embedded (FFPE) tissues with MERSCOPE

Technology Development

POSTER 539

**Jiang He, Justin He, Timothy Wiggin, Rob Foreman, Renchao Chen, Nicholas Fernandez, George Emanuel**

Vizgen, Inc, 61 Moulton St, Cambridge, MA, 02138

FFPE tissues are the most widely used clinical sample types in histology and molecular diagnosis. However, due to RNA degradation and protein crosslinking, FFPE samples are often challenging for single-cell transcriptomic analysis. To accurately profile the gene expression in FFPE samples in situ, a spatial transcriptomics technique with high detection efficiency and single molecule resolution is required. Vizgen's in situ genomics platform MERSCOPE , built on multiplexed error robust in situ hybridization MERFISH technology , directly profiles intact tissue's transcriptome with subcellular spatial resolution. Here, we present a workflow to perform in situ single-cell transcriptomic imaging in FFPE tissue slices and demonstrate MERFISH imaging in more than 10 FFPE sample types from mouse and human, including archival clinical samples. For each tissue type, hundreds of thousands of cells were captured, generating 100s million counts and their spatial information for profiled genes in each sample. Notably, by comparing the performance of MERSCOPE imaging in FFPE and matched fresh frozen samples, we found that the FFPE workflow performs similarly in detection efficiency as compared to the fresh frozen protocol. Furthermore, we demonstrated MERSCOPE FFPE workflow is compatible with protein imaging, and by performing simultaneous protein-based cell boundary staining, MERSCOPE was able to accurately profile gene expression in situ and map cell types in archival clinical human samples. Finally, we constructed a spatially resolved single cell atlas across eight major tumor types, mapped and cataloged different cell types within the tumor microenvironment and systematically characterized the gene expression among cells. Spatially resolved transcriptomic profiling of FFPE samples at single cell level provide enormous opportunities in a wide range of biomedical research areas and will likely open up many applications to study human diseases.

# Standardizing high-throughput, long-read sequencing in 10x Genomics single-cell and spatial applications

Technology Development

POSTER 540

**Carolyn Morrison[1], Juan Pablo Romero[1], Sarah Taylor[1]**

1 10x Genomics, Pleasanton, CA, USA

Alternative transcript isoforms have a range of implications from development to disease. Conventional methods in studying alternative splicing use bulk data from short read sequencers, which gives an average of the transcripts expressed in a cell population but does not provide full length transcript information. In order to unambiguously identify which isoforms are expressed in individual cells, the entire transcript from a barcoded cDNA must be sequenced intact. Here we present two standardized methods for preparing 10x Genomics long read sequencing (LRS) libraries for Oxford Nanopore and PacBio platforms.

cDNA captured using 10x Genomics Single Cell and Spatial Gene Expression assays was partitioned to prepare LRS libraries with Oxford Nanopore's cDNA-PCR kit and PacBio's MAS-ISO-seq kit. These libraries were then sequenced using Oxford Nanopore's PromethION or PacBio's Sequel IIe, respectively. Short read Illumina libraries were also prepared in parallel and sequenced on the NovaSeq. We show that by using either Oxford Nanopore's or PacBio's analysis tools, LRS data is sufficient to perform cell calling and clustering and these results are comparable to the short read data processed through 10x Genomics' Cell Ranger analysis pipeline. LRS data were further analyzed using FLAMES pipeline to perform clustering by isoform gene expression. This analysis enabled the identification of sub-populations of cells that expressed specific isoforms of a gene that was otherwise broadly expressed. For example, we found that expression of HOPX, a factor involved in the terminal differentiation of lung alveolar cells, is also detected in B and T cells in our single cell lung adenocarcinoma dataset, but its isoform expression differs across these two populations. Similarly, we found that Spink8, a serine protease inhibitor detected spatially in the mouse brain hippocampus, predominantly expresses the Spink8-201 isoform.

In summary, we have demonstrated the compatibility of 10x Genomics Single Cell and Spatial Gene Expression assays with Oxford Nanopore and PacBio long read sequencing platforms. By combining 10x assays with long read sequencing technology, one can now identify alternative transcript isoforms with single cell or spatial resolution. We anticipate this area of research will further our understanding of gene regulation in human health and disease.

# Targeted long-read methylation sequencing using nanopore sequencing with hybridization capture



**Masahide Seki[1], Yutaka Suzuki[1]**

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

DNA methylation analysis using long-read sequencing technologies can detect DNA methylation patterns on long DNA molecules and the regions where short reads are hardly aligned. Because most of tissues are composed of several cell species, deep sequencing of the regions regulated by DNA methylation, such as promoters, is more important than shallow sequencing of whole genome to reveal regulation in rare cell types within tissues. However, the existing methods for targeted long-read methylation analysis can process only dozens of target region, and it is impossible to perform methylation analysis of regulome with them.

In this study, we developed targeted nanoEM (t-nanoEM), a method for targeted long-read methylation sequencing by applying hybridization capture to nanoEM, an enzymatic conversion-based long-read methylation analysis method, developed in our previous study. We captured a nanoEM library prepared from 200 ng of genomic DNA with human methylome panel which covers genomic regions over 100 Mb including 4 M of CpG sites. T-nanoEM library was sequenced using a nanopore sequencer PromethION, and we obtained ~3 M of 1d pass reads whose N50 length was ~6 kb. T-nanoEM showed >4-fold higher coverage in target regions, compared with whole genome nanoEM. The methylation rates of t-nanoEM were highly consistent with that of other methods.

# TEM-Seq, an ultrasensitive multiomic platform for epitope-targeted DNA methylation mapping

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

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

Gene expression is regulated by the complex molecular crosstalk between DNA methylation (DNAme) and other chromatin features (e.g. histone post-translational modifications (PTMs) and chromatin-associated proteins (CAPs)). Changes in the chromatin landscape can have a profound impact on DNAme patterning (and vice versa), in both development and disease. However, an understanding of how DNAme co-occurs / co-ordinates with other chromatin features to control gene expression is limited by a lack of reliable genomic tools.

EpiCypher has partnered with New England Biolabs (NEB) to develop Targeted Enzymatic Methylation-sequencing (TEM-seqTM), an ultrasensitive multiomic genomic mapping technology that delivers high-resolution DNAme profiles at epitope-defined chromatin features. TEM-seq marries EpiCypher CUTANA CUT&RUN technology with NEB enzymatic methyl-seq (EM-seq) for unbiased DNAme analysis. EpiCypher is leading the development of ultra-sensitive genomic mapping assays using CUTANATM CUT&RUN to generate truly quantitative data using dramatically reduced cell input and sequencing depth. EM-seq utilizes the enzymatic conversion of DNAme (5mC and 5hmC) and provides a much-needed alternative to bisulfite sequencing (BS; a harsh chemical treatment that introduces DNA breaks and systemic biases) to generate high resolution, unbiased DNAme profiles with 10-fold less sample input (vs. BS).

*(continue to page 208)*

*(continued from page 207)*

To demonstrate assay specificity, sensitivity, and dynamic range we performed the TEM-seq workflow for a diverse set of epigenetic targets in human K562 cells. CUT&RUN steps were controlled by EpiCypher designer nucleosome (dNuc) spike-ins to enable antibody specificity and efficiency monitoring. EM-seq DNAme conversion efficiency was monitored with plasmid DNA spike-ins -/+ 5mC. Of note the dynamic range for DNAme patterns was indistinguishable between 100M and 5M read sequence depths.

We examined TEM-seq multi-omics capabilities in a clinically relevant model. Rett syndrome is a neurodegenerative disorder resulting from mutation of the meCP2 DNAme reader. We created GST-tagged forms of wild-type and mutant meCP2 5mC binding domains and examined their capability on the Luminex platform against fully-defined DNA or Nucleosomes (unmethylated, hemi-methylated or symmetrically-methylated). GST-tagged meCP2 efficiently enriched and recovered 5mC genomic regions. We propose that TEMseq will provide a powerful new approach to advance our understanding of chromatin regulation and unlock the full potential of epigenetic-targeting therapeutics and diagnostics.

# The NovaSeq X and NovaSeq X Plus Sequencing Systems deliver improved sustainability through a broad range of innovations spanning reagent packaging, ambient shipment, and reduced data footprint



**Ros Jackson,** **Valerie Uzzell, Kim Preal, Faith Morel, Alberto Brewer, Michael Chesney, Mike Carney, Carolyn Conant, Johanna Whitacre**

Illumina, Inc., Research and Development, San Diego, CA, USA

The new NovaSeqTM X and NovaSeq X Plus Sequencing Systems deliver multiple sustainability improvements including reagent packaging, ambient shipments, and reduced data footprint. The NovaSeqTM X series packaging leverages optimized design principles to eliminate the need for cushioning insert materials and to consolidate reagent components for reduced shipping and storage space requirements. Multiple components of this new platform have been designed for improve sustainability; secondary carton and tertiary corrugated box materials are made from recycled materials and are also recyclable in a paper recovery stream. This new platform provides a 90% reduction in overall packaging waste; 50% reduction in plastic mass and volume, compared to NovaSeq 6000 Sequencing System. Additionally, this is our first product to include a biopolymer as a primary reagent container. Historically, sensitive reagents and consumables were shipped frozen or refrigerated to provide ideal product quality. Utilizing lyophilized reagents, thermostable components and leveraging data from over 300,000 historical customer shipments we were able to validate our new product to our complex product requirements and global shipping environment by creating an Illumina specific simulated distribution and thermal profile test protocol. After rigorous testing and evaluation, we are pleased to provide a fully ambient shipped product for each customer shipment. In addition to improvements in packaging and shipping, the NovaSeqTM X series has integrated robust data compression algorithms greatly reducing the storage required for high throughput sequencing runs. This data compression achieves up to a 5x reduction in data footprint when compared to the same run without compression, as well as inherently reducing the data transfer energy burden.

# Resolving the exact breakpoints and sequence rearrangements of large neuropsychiatric copy number variations (CNVs) at single base-pair resolution using CRISPR-targeted ultra-long read sequencing (CTLR-Seq)



**Alexander E. Urban**[1], **Bo Zhou**[1], **GiWon Shin**[2], **Lisanne Vervoort**[3], **Stephanie U. Greer**[2], **Yiling Huang**[1], **Tanmoy Roychowdhury**[4], **Reenal Pattni**[1], **Alexej Abyzov**[4], **Joris R. Vermeesch**[3], **Hanlee P. Ji**[2]

1 Stanford University School of Medicine, Department of Psychiatry and Behavioral Sciences, Department of Genetics, Stanford, CA 94305, USA
2 Stanford University School of Medicine, Division of Oncology, Department of Medicine, Stanford, CA 94305, USA
3 KU Leuven, Department of Human Genetics, Leuven, Belgium
4 Mayo Clinic, Department of Health Sciences Research, Rochester, MN 55905, USA

Large recurrent copy number variants (CNVs), i.e. genome aberrations in the form of large deletions and duplications, are by far the highest known risk factors for psychiatric disorders. Highly complex human-specific repetitive sequences on the scale of hundreds of kilobases to megabases i.e. segmental duplications (SegDups) act as substrates that trigger de novo rearrangements resulting such recurrent genome aberrations. These SegDups have so far been impenetrable to direct sequencing analysis. Thus, the exact genomic alterations due to these CNVs and their breakpoint locations remain unknown.

We developed a novel approach (CRISPR-targeted-ultra-long-read sequencing, CTLR-Seq) which combines Cas9-targeting, pulse-field gel electrophoresis, and ultra-long nanopore sequencing to isolate intact large CNV rearrangements in a haplotype-specific fashion to resolve their rearrangement sequences. CTLR-Seq was benchmarked by mapping the breakpoints of the typical 22q11.2 deletion and also applied to the parent of origin where available.

*(continue to page 211)*

*(continued from page 210)*

We applied CTLR-Seq and, for the first time, sequence-resolved the major recurrent CNVs, 16p11.2 deletion (n=4), 16p11.2 duplication (n=8), 15q13.3 deletion (n=4), 15q13.3 duplication (n=3), 1q21 deletion (n=4), 22q11.2 deletion (n=4), mapping out their exact sequence content and rearrangement structure. In 3 of the 4 individual 22q11 patients harboring the typical 3 million bp deletion, we also applied CTLR-Seq to their respective parent of origin to completely resolve the SegDup sequences prior to rearrangement and also for the first time, mapping the typical 22q11 deletion breakpoint in these individuals. These three typical 22q11 deletion breakpoint locations correspond to different retrotransposon (LINE, Alu, and HERV) elements, all within GGT2, suggesting that the mechanism of formation for the typical 22q11.2 deletion is via transposon-mediated recombination and that GGT2 is a recombination hotspot.

CTLR-Seq allows for the complete and haplotype-specific resolution of large recurrent CNVs rearrangements at single base-pair resolution. This will make it possible to analyze cohorts of CNV patients to discover the range of variance within the SegDup rearrangements, breakpoints, and mechanisms of de novo formation, to investigate a possible role of such variance as genetic modifiers, determine the predisposing haplotypes, and to perform functional genomics analyses that include the SegDup sequences and their interactions with the rest of the genome.

# Systematic benchmarking of single-cell ATAC sequencing methods

**Florian De Rop[1]*, Christopher Flerin[1], Paula Soler[2], ..., Satish Pillai[3], Arnau Sebé-Pedrós[2], Leif Ludwig[4], Bart Deplancke[5], Andrew Adey[6], Sarah Teichmann[7], William J. Greenleaf[8], Jason Buenrostro[9], Aviv Regev[10], Holger Heyn[2], Stein Aerts[1]**

1 KU Leuven, VIB Centre for Brain & Disease research, Leuven, Belgium
2 Centro Nacional de Análisis Genómico - Centre de Regulació Genòmica, Barcelona, Spain
3 University of California San Francisco, Department of Laboratory Medicine, San Francisco, CA, USA
4 Max Delbrück Centre, Berlin Institute for Medical Systems Biology, Berlin, Germany
5 École Polytechnique Fédérale de Lausanne (EPFL), Laboratory of Systems Biology and Genetics, Lausanne, Switzerland
6 Oregon Health and Science University, Department of Molecular and Medical Genetics, Portland, OR, USA
7 Wellcome Sanger Institute, Laboratory of Molecular Biology, Cambridge, UK
8 Stanford University School of Medicine, Department of Genetics, Stanford, CA, USA
9 Harvard University, Department of Stem Cell and Regenerative Biology, Cambridge, MA, USA
10 Broad Institute of MIT and Harvard, Klarman Cell Observatory, Cambridge, MA, USA
* Presenting author

Due to sparsity and drop-outs, the analysis and interpretation of scATAC-seq experiments is highly dependent on raw data quality. Several key metrics, such as the number of unique fragments per nucleus, and fragment enrichment in and around peak regions and transcription start sites (TSS), have a major impact on cell filtering, cell-type identification, clustering, and subsequent higher level analyses.

*(continue to page 213)*

*(continued from page 212)*

In order to identify possible technology-specific biases, we performed a large benchmarking study on the entire array of 10x Genomics scATAC-seq methods (v1.0, v1.1, v2.0, multiome and mtscATAC-seq) as well as BioRad ddSeq SureCell ATAC, and two open-source methods (HyDrop-ATAC and s3-ATAC). We performed each technology on a reference sample of peripheral blood mononuclear cells (PBMC), in independent triplicates and on a mixture of two donors where possible. In total, we collected 47 individual PBMC scATAC-seq datasets, executed across 14 institutes. We then developed VSN-ATAC, a universal scATAC data analysis pipeline, to systematically analyse these data. VSN-ATAC accepts raw sequencing data from all 8 scATAC-seq techniques, performs various quality control steps, aligns the sequence reads to the reference genome, and returns a standardised fragments file that can be used for downstream analysis in all current data analysis software packages. We downsampled each dataset to an equal number of reads per cell, and clustered each sample individually using cisTopic. Next, we annotated all cells in each sample using Seurat label transfer from a reference PBMC scRNA-seq dataset. Finally, we identified cell-type specific differentially accessible regions (DARs) and assessed their transcription factor motif enrichment using cisTarget. This standardised analysis workflow allowed us to assess differences in basic quality metrics (number of unique fragments per cell, fraction of reads in peaks, and TSS enrichment scores) as well as higher level quality indicators (label transfer confidence, DAR overlaps and agreement, and motif enrichment scores) between techniques.

While different methods generally agreed on cell type identity and transcription factor activities, they varied widely in library sensitivity, specificity, and higher-level quality indicators. Additionally, our merged dataset totals 168,000 high-quality scATAC-seq profiles and presents a new public reference for chromatin accessibility in human PBMC.

# A cost-efficient workflow for large-scale single-cell RNA-seq using combinatorial indexing and mostly natural sequencing-by-synthesis



**Rahul Satija[1], Longda Jiang[1], Gila Lithwick-Yanai[2], Carol Dalgarno[1], Isabella Mascio[1], Efthymia Papalexi[1], Nika Iremadze[2], Doron Lipson[2]**

1 New York Genome Center, New York, NY
2 Ultima Genomics, 7979 Gateway Blvd, Newark, CA

The ability to generate single-cell RNA-seq datasets at lower cost has been a key driver for large-scale studies such as cell atlases and wide Perturb-seq studies screens involving millions of analyzed cells. In order to fully enable end-to-end cost reduction on a per-cell level, the cost of sample preparation and RNA-seq readout must be reduced in parallel.

Here we benchmark the use of combinatorial cell indexing (Parse Biosciences) together with mostly-natural sequencing-by-synthesis (mnSBS, Ultima Genomics) to efficiently characterize a million cells as part of an on-going Perturb-seq study, in comparison with standard sequencing (Illumina) of the same single-cell libraries.

The analyzed samples include approximately 1 million cells spread across 56 different biological samples, each analyzed in parallel with Perturb-seq. Samples were processed with the Parse Evercode Whole Transcriptome kit, followed by a total of approximately 10B sequencing reads on each sequencing platform. As mnSBS data contains single-ended reads reading through the poly(T) sequence, we developed a method for preprocessing the reads to enable direct analysis with the Parse Biosciences split-seq software.

We benchmark the resulting datasets based on a suite of technical metrics to assess single-cell gene expression. Moreover, we compare the ability of each sequencing technology to detect differentially expressed genes across perturbations, and to group these genes into genetic perturbation response programs. In all cases we obtain highly concordant results, demonstrating the ability to perform massively scalable single-cell perturbation screens with cost-effective library preparation and sequencing approaches.

# BIOSCAN for Flying Insects in the UK Project: Unexpected influences of the UK SARS-CoV-2 surveillance process at the Wellcome Sanger Institute



**Naomi Park[1], Scott Thurston[1], Marco Mosca[1], Emma Betteridge[1], Abdulrahman Tuameh[1], Lyndall Pereira da Conceicoa[1], Mara Lawniczak[1]**

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

The BIOSCAN project will use DNA barcoding to study the genetic diversity of 1,000,000 flying insects from across the UK over a five year period. This project aims to create an improved baseline for the characterisation of insect species diversity and form a foundational resource for DNA-based biomonitoring in the UK.
High-throughput method development requirements at the Wellcome Sanger Institute (WSI) for BIOSCAN DNA barcoding posed the following key challenges:

- Insects undergo non-destructiive extractions without purification prior to amplification, resulting in lysates frequently containing significant amounts of PCR inhibitors
- Any method must be sufficiently sensitive to very low DNA yields from tiny insects
- A degree of normalisation within PCR itself to narrow the range of sequencing coverage between low and high DNA template inputs
- Sufficiently robust to the associated complexities of generating and handling 9216 indexed libaries for pooling onto a single sequencing run
- Compatible with downstream Pacbio and Oxford Nanopore (ONT) long read sequencing, deplexing and upload to mBRAVE for subsequent data analysis

Here we describe how the methodology and processes developed to meet requirements for SARS-CoV-2 surveillance sequencing in the UK have further translated to the Bioscan project, alongside additional optimisations to meet the unique challenges posed by this ambitious project. Through the open sharing of in-house developed protocols we envisage this work will further support the international barcode of lifes global vision to "illuminate biodiversity by developing globally accessible, DNA-based systems for the discovery and identification of all multicellular life."
BIOSCAN - International Barcode of Life (ibol.org)
BIOSCAN – Wellcome Sanger Institute

# Overloading And unpacKing (OAK): A combinatorial indexing method that enables ultra–high-throughput single-cell multiomic profiling



**Bing Wu, Hayley Bennett, Kristel Dorighi, Benjamin Haley, Xin Ye, Zora Modrusan, Spyridon Darmanis**

Genentech Inc, South San Francisco, CA

High-throughput single-cell sequencing is widely used for investigating cellular diversity within complex tissues in health and disease. We developed the 'overloading and unpacking' (OAK) method that allows us to perform single-cell multiomic profiling for a hundred thousand cells or more in a cost-effective, scalable, and experimentally simple manner. OAK utilizes a commonly used droplet indexing system to encapsulate multiple cells in distinct single droplets for the integration of primary barcodes, and releases cells from all droplets for a randomized redistribution of multiple cells into microtiter plates for secondary indexing. This combinatorial indexing strategy largely increases the number of cells that can fit in a single run of droplet generation, while featuring fast and convenient wet lab handling. The resulting libraries remain divisible to allow flexible selection of cell numbers for sequencing. We demonstrated that OAK is readily compatible with a variety of applications, such as transcriptome profiling, RNAseq coupled with ATACseq, and sample multiplexing. When implemented in single-cell CRISPR screens, the OAK approach efficiently identifies candidate mutations and transcriptome changes that confer resistance to drug treatment in cancer cell lines.

# Multiplexed single-cell surface protein detection and RNA profiling of fixed cells



**Xue Wu[1], Paul Lund[1], Ryan Stott[1], Shaun Jackman[1], Luiz Irber[1], Patrick Rolli[1], Kevin Chen[1], Peter Smibert[1], Andrew Kohlway[1]**

10x Genomics, Inc., Pleasanton, CA, USA

With recent advances in technology, single cell profiling has entered the multiomics age. More researchers are combining single cell gene expression with additional sources of information to obtain a more comprehensive understanding of cellular events. The 10x Genomics Chromium Fixed RNA Profiling (FRP) assay provides single cell whole transcriptome RNA profiling in formaldehyde-fixed cells and is compatible with Feature Barcode technology for antibody-based cell surface protein detection using BioLegend TotalSeq™-B antibodies. However, the current built-in barcoding approach that allows for sample multiplexing can only be applied to RNA profiling, thus limiting its utility in large-scale single cell multiomics studies. We developed an FRP-compatible workflow to enable sample multiplexing for cell surface protein detection in cells stained with BioLegend TotalSeq™-C antibodies. Following antibody staining and fixation, sample-specific barcode oligos with the FRP capture sequence and a splint compatible with the TotalSeq™-C capture sequence are added to the overnight gene expression (GEX) probe hybridization. After GEM generation, the multiplexing barcode oligo is first splint-ligated onto the antibody conjugated Feature Barcode oligos and then extended onto the GEM barcode oligo concurrently with GEX probe ligation and barcoding.

The addition of a second barcode to the antibody Feature Barcode oligo upstream of GEMs allows the antibody libraries to be demultiplexed together with the FRP GEX libraries. Detection of TotalSeq™-C antibodies in the Chromium FRP workflow was validated by co-staining samples with both TotalSeq™-B and -C antibodies and comparing the TotalSeq™-C antibody signal in the multiplexable “indirect-capture” workflow to the TotalSeq™-B antibody signal in the currently supported “direct-capture” workflow.

*(continue to page 218)*

*(continued from page 217)*

We demonstrated the utility of this sample multiplexing approach on peripheral blood mononuclear cells stained with BioLegend TotalSeq™-C Human Universal Cocktail with a 4 and 16 sample multiplexing configurations. This new method unleashes the full sample multiplexing capabilities of our Chromium Fixed RNA Profiling kit for both RNA profiling and cell surface protein detection, and further expands the antibody panels that can be used in the Fixed RNA Profiling workflow. We anticipate that this additional feature will enable broader adoption of Chromium Fixed RNA Profiling in larger scale multiomics studies involving fixed samples.


Parse Biosciences Evercode™ single-cell technology enables all scales of experiments with unmatched sensitivity and without the limitations of today's microfluidics approaches.

## Unlock your lab's potential in suite 215.

Parse BIOSCIENCES

# AGBT2023™

## PRECISION HEALTH


**SEPTEMBER 7–9, 2023**

HILTON BAYFRONT HOTEL
SAN DIEGO, CA

REGISTRATION AND
CALL FOR ABSTRACTS
OPENS THIS MAY

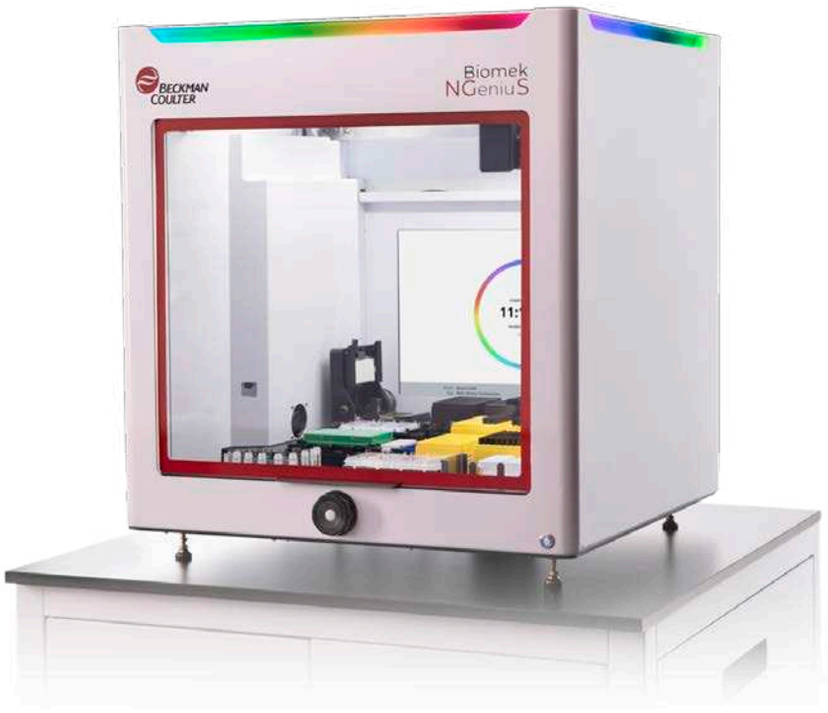
LEARN MORE AGBT.ORG

# STRIKE WORKFLOW GOLD

## JOIN US IN SUITE 320

The Biomek NGeniuS Next Generation Library Prep System is an evolutionary leap in NGS liquid handling. With it in your lab, you can:

- Easily set up an application and complete a sample run
- Automate error-prone manual processes and pipetting
- Reduce hands-on time
- Run an array of kits from multiple vendors




**LEARN MORE**

BECKMAN COULTER Life Sciences



Biomek NGeniuS Next Generation Library Prep Systems and demonstrated applications are not intended or validated for use in the diagnosis of disease or other conditions.

For Beckman Coulter's worldwide office locations and phone numbers, please visit Contact Us at beckman.com

2022-AMER-EN-100566-v1

# Index

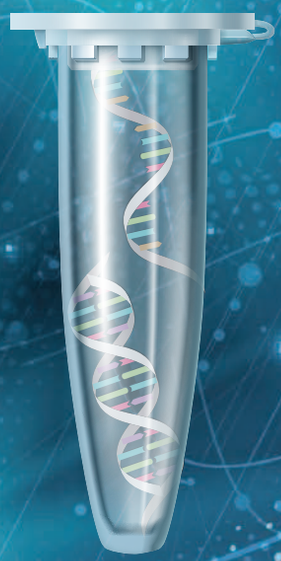
| | | | | |
|---|---|---|---|---|
| E | Ebenstein | Yuval | Tel-Aviv University | 53 |
| | Evanich | Daniel | New England Biolabs | 123 |
| F | Fan | Li | Genmab | 37 |
| | Faria de Oliveira | Michelli | 10xGenomics | 169 |
| | Fredriksson | Simon | PixelgenTechnologiesAB | 202 |
| G | Gans | Joseph | Cellarity | 137 |
| | Garcia Medina | Sebastian | Weill Cornell Medicine | 107 |
| | Garcia-Bassets | Ivan | Universal Sequencing Technology Corp | 153 |
| | Geiger | Heather | New York Genome Center | 82 |
| | Geiss | Gary | University of Wisconsin School of Medicine & Public Health | 36 |
| | Gillespie | Meghan | Broad Institute of MIT and Harvard | 127 |
| | Greer | Stephanie | MyOme, Inc. | 155 |
| | Grills | George | University of Miami Miller School of Medicine | 136 |
| | Grimes | Susan | Stanford University School of Medicine | 38 |
| | Goodwin | Sara | Cold Spring Harbor Laboratory | 151 |
| H | Hamilton | Gordon | Depixus SAS | 180 |
| | He | Jiang | Vizgen, Inc. | 204 |
| | Heider | Margaret | New England Biolabs Inc. | 25 |
| | Hoang | Margaret | NanoString® Technologies Inc. | 170 |
| | Höijer | Ida | Uppsala University | 128 |
| | Hook | Paul | Johns Hopkins University | 193 |
| | Hwang | William | Massachusetts General Hospital | 41 |
| I | Iyer | Shruti | Cold Spring Harbor Laboratory | 173 |
| J | Jackson | Ros | Illumina, Inc. | 209 |
| | Jason | Navarro | Institute for Genomic Medicine | 146 |
| | Jenike | Katharine | Computer Science, JHU, | 94 |

Join us for our talk

Roche

# Single-tube total nucleic acid library preparation for FFPE:

## *A combined method for DNA and RNA*



## Visit us at booth 319 to learn more about:

**Custom library preparation solutions:**
Pick and choose the best combinations of KAPA library prep reagents to optimize your workflows and minimize waste.

**Primer Extension Target Enrichment with KAPA HyperPETE:**
Achieve superior performance and coverage uniformity compared to hybridization-based workflows—within a single day.

**Digital LightCycler® dPCR System:**
Use one powerful clinical research tool to achieve sensitivity, precision, flexibility, and integration.

Learn more at: **go.roche.com/AGBT2023**

**February 8, 2023**
4:05 pm – 4:20 pm
Grand Ballroom

**Speaker:**

**Mariana Kiehl, PhD.**
Senior Applications Scientist
Roche Sequencing & Life Science



MC-US-12337 A719 12/22

WATCHMAKER GENOMICS

**JOIN US** | WEDNESDAY, 4:20 PM | GRAND BALLROOM

# ADVANCING NGS LIBRARY PREPARATION USING CUTTING-EDGE ENZYME ENGINEERING APPROACHES

Come learn about Watchmaker's unique innovation platform that layers specifically engineered enzymes into rapid NGS library prep solutions tailored for increased sensitivity, including our:

Click to visit our site and learn more about our solutions!

**Watchmaker RNA Library Prep Kit with Polaris Depletion** that generates sequencing-ready libraries in under 4.5 hours and improves performance with challenging samples

**Watchmaker DNA Library Prep Kit** that combines scalability and accuracy—reducing sequence artifacts by up to 90%

**Equinox Library Amplification Kit**, which reduces misincorporation rates by up to 40% to enable highly sensitive applications


**Stop by Suite 314** and pick up the coolest swag, including our new (*and uber addictive*) galaxy cube!

watchmakergenomics.com/AGBT23

SINGULAR GENOMICS

# Discover G4

Explore how the G4 Sequencing Platform can deliver you fast, accurate results with more flexibility today in Suite 312.

## POWER

**15-400 Gb Output**

Delivers more data per day than any benchtop sequencer.

## SPEED

**6-19 Hour Run Times**

Raising the bar with industry-leading run times.

## FLEXIBILITY

**4 Flow Cells 16 Lanes**

Intentionally designed to maximize operational efficiency.

Learn More

Singulargenomics.com