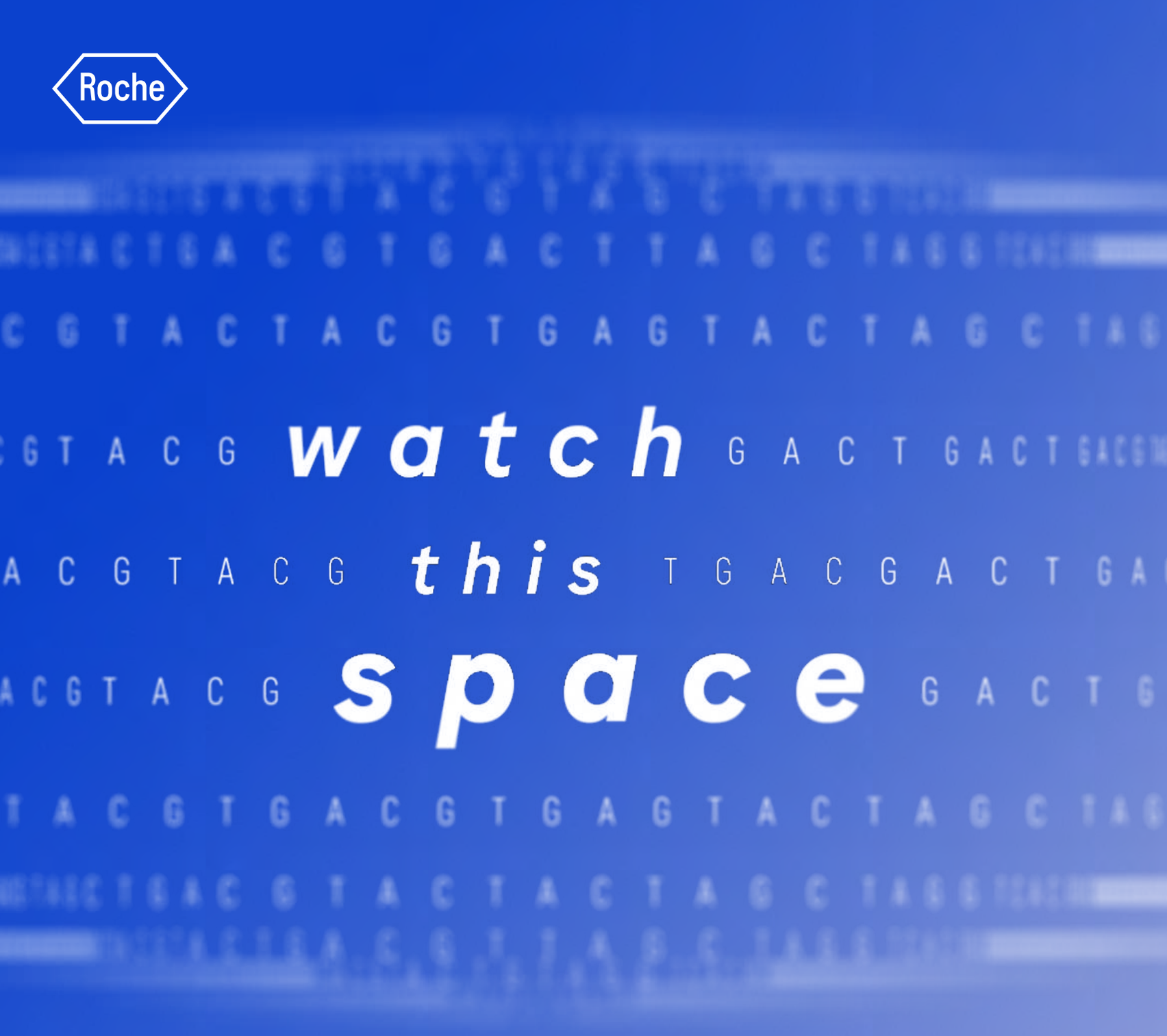
25 YEARS OF AGBT™

February 23-26, 2025 • Marco Island, FL



Ready to see what's next in NGS?

Roche Sequencing Solutions is introducing the future of NGS technology.

Join us to hear how our novel technology could expand your sequencing goals.

Head to Osprey 2 for exclusive in-suite events as we make space for your advances.

Welcome to the 25th Anniversary Advances in Genome Biology and Technology (AGBT) General Meeting!

We are thrilled to celebrate this milestone event with you as AGBT marks 25 years of fostering collaboration, innovation, and progress in the genomic research community.

Over the next few days, we'll delve into groundbreaking advancements in DNA sequencing technologies, cutting-edge experimental and analytical approaches, and their transformative applications across diverse fields. This meeting serves as a unique platform to showcase technological innovation, foster interdisciplinary collaboration, and inspire groundbreaking discoveries, bringing together researchers, scientists, and visionaries from around the globe.

As part of our commitment to advancing genomic science and technology, AGBT is proud to host three distinguished conferences and networking events each year. Save the dates for our upcoming events:

AGBT Agriculture: *Innovating Agriculture through Genomics*
March 31–April 2, 2025, in Orlando, Florida (Registration open now)

AGBT Precision Health: *Exploring AI, Rare Diseases, and Transformative Genomics*
September 8–10, 2025, in San Diego, California

Thank you for joining us as we commemorate 25 years of discovery and excellence. Let's make this a truly memorable meeting!

AGBT is a not-for-profit organization. Our mission is to advance research, promote education, and expand commerce in genome science and technology.

AGBT2025™
AGRICULTURE

MARCH 31 – APRIL 2, 2025
CARIBE ROYALE HOTEL | ORLANDO, FL

REGISTRATION NOW OPEN | LEARN MORE AT AGBT.ORG

Organizing Committee

We are incredibly grateful to our meeting Co-Chairs Eric Green, Len Pennacchio, Beth Shapiro, and the entire Scientific Organizing Committee for the countless hours they invested in making AGBT25 a content-rich experience.

Penelope Bonnen
Baylor College of Medicine

Federica Di-Palma
Genome British Columbia

Kim Doheny
Johns Hopkins University School of Medicine

Stacey Gabriel
Broad Institute of MIT and Harvard

***Eric Green**
National Human Genome Research Institute (NHGRI)

Elinor Karlsson
Broad Institute of MIT and Harvard, UMass Chan Medical School

Chris Mason
Weill Cornell Medicine

John McPherson
University of California, Davis

Stephen Montgomery
Stanford University

***Len Pennacchio**
Lawrence Berkeley National Lab/Joint Genome Institute

Michael Schatz
Johns Hopkins University

***Beth Shapiro**
University California, Santa Cruz

Mike Zody
New York Genome Center

**Meeting Co-Chairs*

We gratefully acknowledge the generous support provided by the following companies

GOLD

illumina®

SILVER

Roche

ULTIMA GENOMICS

BRUKER

BRONZE

WATCHMAKER GENOMICS

IDT INTEGRATED DNA TECHNOLOGIES

New England Biolabs

Complete GENOMICS

Olink Part of Thermo Fisher Scientific

PARSE BIOSCIENCES

SCALE biosciences

PacBio

CONTRIBUTING SPONSORS

10x GENOMICS

Agilent

AlphaCyzyme part of Maravai LifeSciences

CHRYSALIS BIOMEDICAL ADVISORS

Codetta Bio

Element Biosciences

Genome Research

GINKGO

INOCRAS

lifebit

MICRONBRANE MEDICAL

Moleculent

Nabsys

Oxford Nanopore Technologies

PIXELGEN TECHNOLOGIES

QIAGEN

revvity

SINGULAR GENOMICS

Stellaromics

TD Cowen a division of TD Securities

TWIST BIOSCIENCE

Vazyme

vizgen

Volta Labs

General Information

Registration / Hospitality Desk

Saturday, February 22	12:00 p.m. – 8:00 p.m.
Sunday, February 23	8:00 a.m. – 8:00 p.m.
Monday, February 24	7:30 a.m. – 9:30 p.m.
Tuesday, February 25	7:30 a.m. – 8:00 p.m.
Wednesday, February 26	7:30 a.m. – 6:00 p.m.

Name Badges

Badges are essential for conference security and networking. All attendees must:

-Always wear badges visibly, clipped to the provided lanyard with the name facing outward
-Keep badges with you throughout the event
-Present badges for entrance to conference entrances and events

Important: Badges are non-transferable. Proper badge display ensures attendee safety and facilitates valuable connections during our 25th Anniversary meeting.

Transportation

Airport Shuttle Service:

AGBT is pleased to offer complimentary shuttle transportation between Southwest Florida International Airport (RSW) and the JW Marriott Marco Island (48 miles, approximately 1-hour transfer).

-Arrival Shuttles (every hour):
  -Pick-up location: Baggage Claim Area
    -Saturday, February 22: 10:00 a.m. - 10:00 p.m.
    -Sunday, February 23: 10:00 a.m. - 8:00 p.m.

-Departure Shuttles (every 30 minutes):
  -Pick-up location: San Marco Entrance
    -Thursday, February 27: 3:30 a.m. - 4:00 p.m.

Upon arriving at Southwest Florida International Airport, go to the baggage claim area, even if you didn't check luggage. A uniformed NT&T Airport Coordinator, identifiable by a red golf shirt and black slacks, will greet you with an AGBT sign. If you have trouble finding the greeter, contact Naples Transportation & Tours (NT&T) at 239-260-3262.

Alternative Transportation Options:

-To arrange a private transfer, please click this link
-Taxi Estimated Cost: $85-$110 one-way
-Rideshare Services: Uber/Lyft available (estimated $75-$95)
-Rental Car: Several agencies located at the airport
-Private Vehicle: Detailed driving directions available through the hotel concierge desk

Social Events

Welcome Reception: Sunday, February 23, 7:00 p.m. – 10:00 p.m., Resort Beach
Women's Networking Event: Monday, February 24, 5:15 p.m. – 7:15 p.m., Calusa Terrace
Sponsor's Social (Connect, Collaborate, and Chill): Monday, February 24, beginning at 9:30 p.m., Osprey rooms, Caxambus, Sirene 5th Floor rooms, Sirene tower and Lanai Suites, Presidential and Terrace Suites
AGBT Dinner: Tuesday, February 25, 6:10 p.m. – 7:25 p.m., Sunset Terrace
Farewell Dinner: Wednesday, February 26, 7:00 p.m. – Midnight, Banyan Ballroom

Conference Technology

Enhance your conference experience with our comprehensive mobile app, designed to keep you connected throughout the 25th Anniversary meeting.
App Features:

-Interactive conference agenda
-Detailed speaker and abstract information
-Networking tools
-Interactive venue map
-Real-time schedule updates

Download Instructions:

-Look for an email closer to the meeting with more information on how to download the mobile app
-Download app from the AGBT email link
-iOS: Scan QR code or visit App Store
-Android: Scan QR code or visit Google Play
-Alternative: Scan QR code at the registration desk or ask for assistance

Official AGBT Mobile App

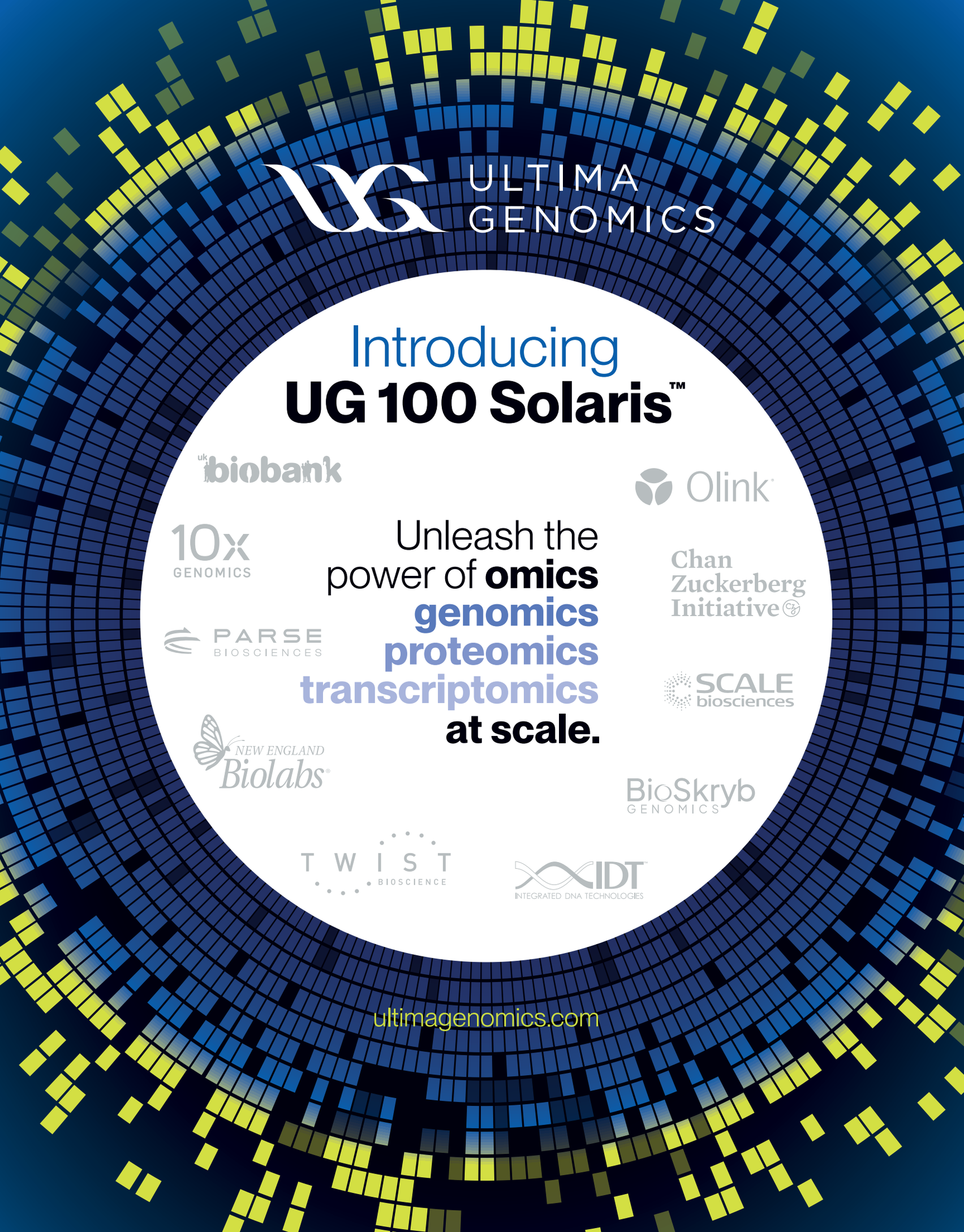
Stay connected and share your AGBT experience!

-Primary Conference Hashtag: #AGBTGM
-Anniversary-Specific Hashtag: #25YearsOfAGBT

Tech Support:

-AGBT Hospitality Desk: Calusa Foyer
-Wi-Fi Network: **AGBTGM2025**
-Wi-Fi Password: **NextStopAGBTAg**

Capture the moment, connect with colleagues, and celebrate 25 years of scientific advancement!


ULTIMA GENOMICS
Introducing
UG 100 Solaris™
uk biobank
Olink
10x GENOMICS
Unleash the power of omics
genomics
proteomics
transcriptomics
at scale.
Chan Zuckerberg Initiative
PARSE BIOSCIENCES
SCALE biosciences
NEW ENGLAND Biolabs
BioSkryb GENOMICS
TWIST BIOSCIENCE
IDT INTEGRATED DNA TECHNOLOGIES
ultimagenomics.com

Agenda

Saturday, February 22

10:00 a.m. – 10:00 p.m.	Transfers from Southwest Florida International Airport (RSW)
12:00 p.m. – 8:00 p.m.	**Meeting Registration / Hospitality Desk** Calusa Foyer

Sunday, February 23

8:00 a.m. – 8:00 p.m.	**Meeting Registration / Hospitality Desk** Calusa Foyer
10:00 a.m. – 8:00 p.m.	Transfers from Southwest Florida International Airport (RSW)
1:00 p.m. – 4:00 p.m.	**10x Genomics Pre-Conference** “Revving Up Insights in Single Cell and Spatial Explorations: The 10x 2025 Roadmap” Banyan Ballroom

Scientific Program

Plenary Session:	**Genetics (Len Pennacchio, Lawrence Berkeley National Laboratory and DOE Joint Genome Institute, Session Chair)** Calusa Ballroom 1-7
5:00 p.m. – 5:10 p.m.	**Opening Remarks (Richard K. Wilson, founding executive director, Institute for Genomic Medicine at Nationwide Children’s Hospital, the Nationwide Foundation Endowed Session Chair in Genomic Medicine, and professor of Pediatrics at The Ohio State University College of Medicine)**
5:10 p.m. – 5:40 p.m.	**David Altshuler, Vertex Pharmaceuticals** “Humanizing drug discovery”
5:40 p.m. – 6:10 p.m.	**Ewan Birney, European Molecular Biology Laboratory (EMBL)** "Genomics - looking back, looking forwards. A view from EMBL"

6:10 p.m. – 6:30 p.m.	**Dewi Moonen (Abstract Selected Talk) European Molecular Biology Laboratory (EMBL)** "A large scale CD4+ T-cell enhancer screen links immune disease variants to target genes"
6:30 p.m. – 6:50 p.m.	**Orly Alter, (Abstract Selected Talk), University of Utah and Prism AI Therapeutics, Inc.** "Multi-tensor AI/ML uniquely able to identify actionable and mechanistically interpretable predictors from small-cohort and noisy high-dimensional multi-omic clinical data"
7:00 p.m. – 10:00 p.m.	**Welcome Reception** Resort Beach
Monday, February 24	**(Morning)**
7:30 a.m. – 9:00 a.m.	**Breakfast Sponsored by Roche** Sunset Terrace
7:30 a.m. – 9:30 p.m.	**Meeting Registration / Hospitality Desk** Calusa Foyer
Plenary Session:	**Genomics I (Federica Di Palma, Genome British Columbia, Session Chair)** Calusa Ballroom 1-7
9:00 a.m. – 9:30 a.m.	**Richard Lifton, Rockefeller University** "Genetic Contributions to Extreme Human Phenotypes and Their Implications"
9:30 a.m. – 10:00 a.m.	**Tuuli Lappalainen, New York Genome Center, Columbia University** "Functional variation in the genome: lessons from natural and experimental perturbations"
10:00 a.m. – 10:30 a.m.	**Richard Gibbs, Baylor College of Medicine** "Somatic Mutation in Rare and Common Disease"

10:30 a.m. – 11:00 a.m.	**Coffee Break Sponsored by Revvity** Sponsor Promenade – Osprey Foyer
11:05 a.m. – 11:25 a.m.	**Tim Coorens (Abstract Selected Talk), Broad Institute of MIT and Harvard** “Novel technologies to detect somatic mutations across normal tissues”
11:25 a.m. – 11:45 a.m.	**Poster Flash Talks**
Monday, February 24	**(Afternoon)**
12:00 p.m. – 1:30 p.m.	**Gold Sponsor – Illumina – Workshop & Lunch - Cande Rogert, PhD, Vice President, Global Head of Advanced Sciences, Illumina and Sergio Peisajovich, PhD, Vice President, R&D, Illumina** “Driving the Multiomics Revolution: The Latest Innovations from Illumina” Banyan Ballroom
12:00 p.m. – 1:30 p.m.	**AGBT Lunch** Sunset Terrace, Lanai Beach
1:30 p.m. – 3:30 p.m.	**Poster Session with Coffee & Dessert** Calusa Ballroom 8-12
Plenary Session:	**Biology (John McPherson, University of California, Davis, Session Chair)** Calusa Ballroom 1-7
3:30 p.m. - 4:00 p.m.	**Krystal Tsosie, Assistant Professor, School of Life Sciences, Arizona State University** "Building data trusts and trust in data: Indigenous-led genomic databases for health equity futures"
4:00 p.m. – 4:30 p.m.	**Corrie Painter, Broad Institute of MIT and Harvard** “No Data, No Limits: Innovating Through Adversity for Rare Cancers and Beyond”
4:30 p.m. – 4:50 p.m.	**Stephen Kingsmore, (Abstract Selected Talk) Rady Children’s Institute for Genomic Medicine** “A novel WGS workflow to aid in the design of allele-specific antisense oligonucleotides for patients with rare genetic conditions”

4:50 p.m. – 5:10 p.m.	**Samouil L. Farhi (Abstract Selected Talk), Broad Institute of MIT and Harvard** "Simultaneous CRISPR screening and spatial transcriptomics reveals intracellular, intercellular, and functional transcriptional circuits"
Monday, February 24	**(Evening)**
5:15 p.m. - 7:15 p.m	**Dinner on Your Own**
5:15 p.m. – 7:15 p.m.	**Women's Networking Event** Empowering Talk by Tuuli Lappalainen Calusa Terrace
Concurrent Session:	
7:30 p.m. - 9:30 p.m.	**Cancer (Kim Doheny, Johns Hopkins University School of Medicine, Session Chair)** Banyan Ballroom
7:30 p.m. – 7:50 p.m.	**Omer Bayraktar, Wellcome Sanger Institute** "Mapping the rules of glioblastoma with integrated single cell and spatial genomics"
7:50 p.m. – 8:10 p.m.	**Stephanie Chrysanthou, Memorial Sloan Kettering Cancer Center** "Detection of germline alterations in cancer patients by nanopore adaptive sampling"
8:10 p.m. - 8:30 p.m.	**William Hwang, Massachusetts General Hospital / Harvard Medical School** "Whole transcriptome spatial molecular imaging in matched pre and post treatment specimens to identify therapeutic vulnerabilities and treatment-associated effects in pancreatic adenocarcinoma"
8:30 p.m. – 8:50 p.m.	**Mengxiao He, SciLifeLab, Kungliga Tekniska Högskolan (KTH)** "Comprehensive spatial profiling of prostate cancer metastasis: Mapping clonal evolution, microenvironment dynamics, and early metastatic markers"

8:50 p.m. – 9:10 p.m.	**Corinne Strawser, Nationwide Children's Hospital** "Integrative spatial analysis to investigate the impact of oncolytic herpes simplex virus treatment on the glioblastoma microenvironment"
9:10 p.m. - 9:30 p.m.	**Ulf Gyllensten, SciLifeLab, Uppsala University** "Precision diagnosis of ovarian cancer – Identification of an eight-plasma protein biomarker panel that separate benign from malignant tumors"

Concurrent Session:

7:30 p.m. – 9:30 p.m.	**Computational Biology (Elinor Karlsson, Broad Institute of MIT and Harvard, UMass Chan Medical School, Session Chair)** Calusa Ballroom 1-7
7:30 p.m. – 7:50 p.m.	**Konrad Karczewski, Massachusetts General Hospital/Broad Institute** "Decoding the impact cascade using deep learning: from variant effects on genes to genic loss consequences in humans"
7:50 p.m. – 8:10 p.m.	**Xuefang Zhao, Massachusetts General Hospital** "A global reference of 2.7M structural variants and their functional prediction across gnomAD and All of Us"
8:10 p.m. – 8:30 p.m.	**Alexander Lucaci, Weill Cornell Medicine** "Diversity and distinctive traits of the global RNA virome in urban environments"
8:30 p.m. – 8:50 p.m.	**Ioannis Mouratidis, Penn State College of Medicine** "Nucleic Quasi-Primes: Identification of the Shortest Unique Oligonucleotide Sequences in a Species"
8:50 p.m. – 9:10 p.m.	**Hongyu Zhao, Yale University** "INSPIRE: interpretable, flexible and spatially-aware integration of multiple spatial transcriptomics datasets from diverse sources"
9:10 p.m. – 9:30 p.m.	**Nicholas Tatonetti, Cedars-Sinai Medical Center** "Running genome-wide association studies is less than five seconds"

Starting at 9:30 p.m.	**Sponsor's Social (Connect, Collaborate, and Chill)** Sirene Tower Suites, Sponsor Promenade, Caxambus
Tuesday, February 25	**(Morning)**
7:30 a.m. – 9:00 a.m.	**Breakfast Sponsored by Pixelgen Technologies** Sunset Terrace
7:30 a.m. – 8:00 p.m.	**Hospitality Desk** Calusa Foyer
Plenary Session:	**Genomics II (Mike Zody, New York Genome Center, Session Chair)** Calusa Ballroom 1-7 Ballroom
9:00 a.m. – 9:30 a.m.	**Jeff Gordon, Washington University** "Developing microbiome-targeted therapeutics for childhood undernutrition"
9:30 a.m. – 10:00 a.m.	**Christina Warinner, Harvard University** "The archaeology of microbes and the future of lost genomes"
10:00 a.m. – 10:20 a.m.	**Tracy Bedrosian (Abstract-Selected Talk), Institute for Genomic Medicine at Nationwide Children's Hospital** "Profiling somatic variants in human brain tissue with single-cell transcriptomics"
10:20 a.m. – 11:00 a.m.	**Coffee Break Sponsored by Roche** Sponsor Promenade – Osprey Foyer
11:00 a.m. – 11:20 a.m.	**Athma Pai, (Abstract Selected Talk) University of Massachusetts, Chan Medical School** "mRNA transcription and termination are spatially coordinated"
11:20 a.m. – 11:40 a.m.	**Next Gen Leadership Awards (Elinor Karlsson, Broad Institute of MIT and Harvard, UMass Chan Medical School, Beth Shapiro, University California, Santa Cruz, Session Co-Chairs**

Tuesday, February 25	**(Afternoon)**
12:00 p.m. – 2:15 p.m.	**Silver Sponsor Workshops and Lunch** Banyan Ballroom
12:05 p.m. – 1:05 p.m.	**Silver 1 Sponsor Workshop and Lunch - Roche** Join Roche experts and early evaluators to hear how SBX is making space for new advances
1:10 p.m. – 2:10 p.m.	**Silver 2 Sponsor Workshop and Lunch - Ultima Genomics** "Beyond the $100 genome – population-scale proteomics, billions of cells for AI/ML, ppm cfDNA accuracy, and more"
12:00 p.m. – 2:10 p.m.	**AGBT Lunch** Sunset Terrace, Lanai Beach
2:15 p.m. – 4:44 p.m.	**Bronze Sponsor Workshops (Stephen Montgomery, Stanford University, Session Chair)** Calusa Ballroom 1-7
2:20 p.m. – 2:40 p.m.	**Bronze 1 Sponsor – Watchmaker Genomics - Craig Marshall, Senior Research Scientist** "The Power of Positive: TAPS Methylation Analysis Unlocks Comprehensive Tumor Multiomics"
2:40 p.m. – 3:00 p.m.	**Bronze 2 Sponsor – Integrated DNA Technologies – Steven Henck, VP of NGS, Fellow, Danaher- Life Sciences Division** "NGS-Inspired Innovations: Driving the Future of Gene Reading"
3:00 p.m. – 3:20 p.m.	**Bronze 3 Sponsor – New England Biolabs, Inc. - Ivo Glynne Gut, Director, Centro Nacional de Analisis Genomico, Barcelona, Spain** "Epigenetic Analyses Then and Now"
3:20 p.m. – 3:35 p.m.	**Bronze 4 Sponsor – Complete Genomics - Piotr Mieczkowski, MSc, PhD, Professor & Director of the High Throughput Genomic Sequencing Facility, Genetics University of North Carolina** "Advancing Genomic Discoveries with the Latest DNBSEQ Sequencer"

3:35 p.m. – 3:50 p.m.	**Bronze 5 Sponsor – Olink (part of ThermoFisher) - Cindy Lawley Sr Director Population Health at Olink part of ThermoFisher & Christopher E. Mason Professor of Genomics, Physiology, and Biophysics at Weill Cornell Medicine** "Breaking Barriers in Proteomics with Olink's new scalable NGS Based Platform for Transformative Research"
3:50 p.m. – 4:05 p.m.	**Bronze 6 Sponsor - Pranav Patel, Co-Founder and CEO - iconPCR, n6** "iconPCR from n6: The Next Frontier in Next Generation Sequencing"
4:05 p.m. – 4:20 p.m.	**Bronze 7 Sponsor – Parse Biosciences - Charlie Roco, PhD, Chief Technology Officer at Parse Biosciences** "Ushering in a new era of biological discovery with scaled up single cell sequencing"
4:20 p.m. – 4:32 p.m.	**Bronze 8 Sponsor - Scale Biosciences - Giovanna Prout, CEO** "Defy single cell gravity: Escape platform constraints in single cell"
4:32 p.m. – 4:44 p.m.	**Bronze 9 Sponsor - PacBio - Melissa Smith, PhD, Associate Professor, Department of Biochemistry & Molecular Genetics, University of Louisville, and CEO of Clareo** "Decoding Diversity: How HiFi Sequencing is Revolutionizing Immune Response and Vaccine Innovation"
4:45 p.m. – 6:10 p.m.	**Poster Session and Wine & Cheese Reception Sponsored by Roche** Calusa Ballroom 8-12
Tuesday, February 25	**(Evening)**
6:10 p.m. – 7:25 p.m.	**AGBT Dinner** Sunset Terrace, Lanai Beach
Concurrent Sessions:	
7:30 p.m. – 9:30 p.m.	**Technology (Chris Mason, Weill Cornell Medicine, Session Chair)** Calusa Ballroom 1-7
7:30 p.m. – 7:50 p.m.	**Michal Lipinski, Broad Institute of MIT and Harvard** "Large continuous area of spatial transcriptome profiling allows for efficient tissue characterization with high resolution and limited batch effect"

7:50 p.m. – 8:10 p.m.	**Katherine Miller, Nationwide Children's Hospital** "Development of personalized liquid biopsy assays for pediatric solid tumors"
8:10 p.m. – 8:30 p.m.	**Bo Zhou, Stanford University School of Medicine** "Detection and analysis of complex structural genome variation across human populations and in brains of individuals with psychiatric disorders"
8:30 p.m. – 8:50 p.m.	**Darren Segale, Illumina** "A high-resolution spatial transcriptomic map of the pregnant mouse brain reveals regionally distinct gene expression regulation related to maternal behavior"
8:50 p.m. – 9:10 p.m.	**Joseph Beechem, Bruker** "Sub-cellular, multiomic imaging of the entire protein-coding transcriptome (18936-plex RNA) and 75-plex Protein on a single-slide FFPE tissue section using CosMx Spatial Molecular Imaging (SMI)"
9:10 p.m. – 9:30 p.m.	**Ronan Chaligne, Memorial Sloan Kettering Cancer Center** "Transcript-specific enrichment enables profiling rare cell states via single-cell RNA sequencing"

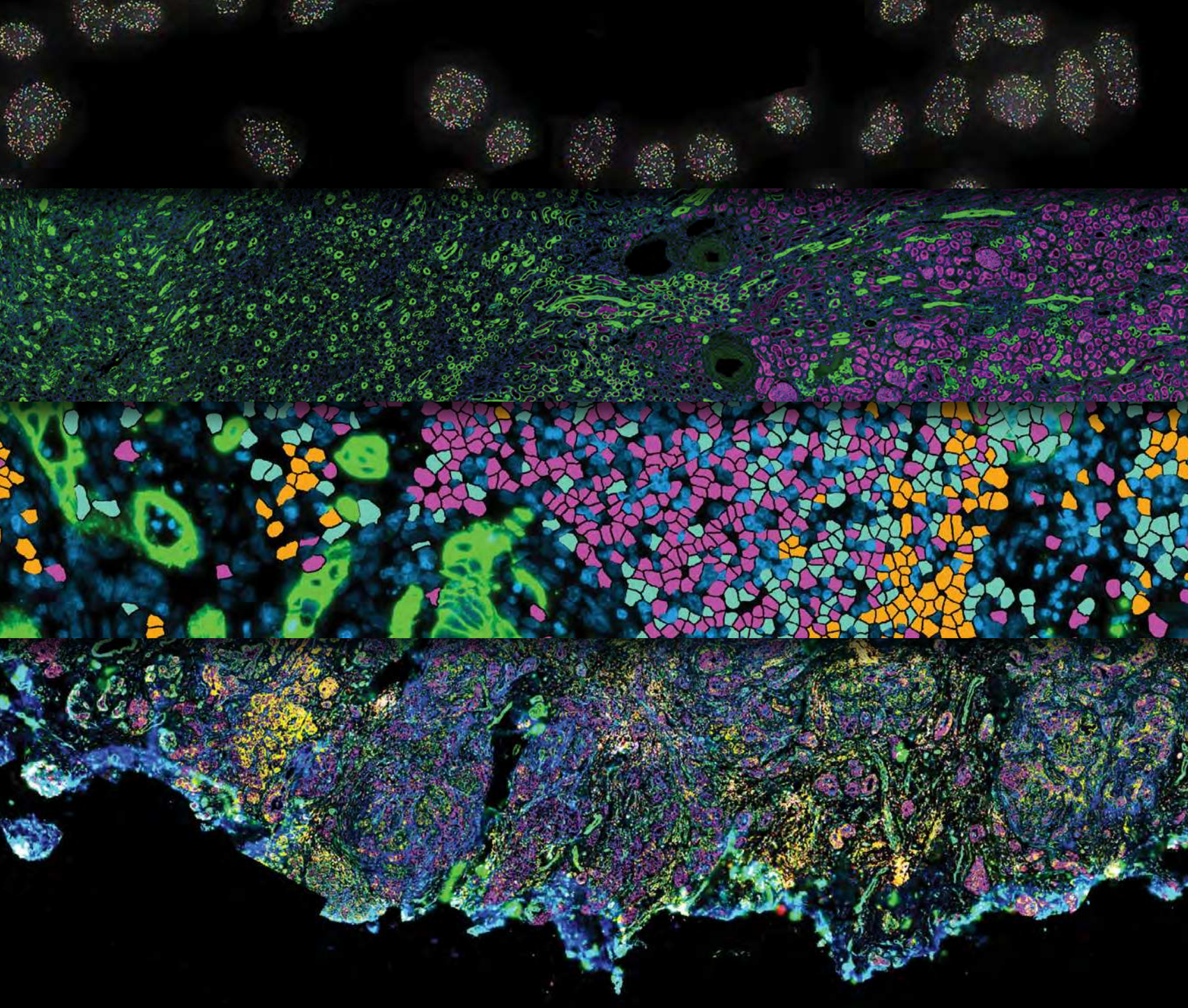
Concurrent Sessions:

7:30 p.m. – 9:30 p.m.	**Biology (Penelope Bonnen, Baylor College of Medicine, Session Chair)** Banyan Ballroom
7:30 p.m. – 7:50 p.m.	**Tamas Korcsmaros, Imperial College London** "Decoding non-coding SNPs – Patient-specific regulatory network rewiring in inflammatory bowel disease unravels how genetic polymorphisms divert incoming environmental signals, contribute to disease phenotypes and stratify patients"
7:50 p.m. – 8:10 p.m.	**Dalila Pinto, Icahn School of Medicine at Mount Sinai** "Long-read RNA Isoform map of the Human Brain (IsoHuB)"

8:10 p.m. – 8:30 p.m.	**Francisco De La Vega, Galatea Bio, Inc.** "Uncovering Genomic Diversity in the African Diaspora via Long-Read Sequencing"
8:30 p.m. – 8:50 p.m.	**Ava Hoffman, Fred Hutch Cancer Center** "BioDIGS: Illuminating soils with BioDiversity and Informatics for Genomics Scholars"
8:50 p.m. – 9:10 p.m.	**Timothy Smith, USDA, ARS, U.S. Meat Animal Research Center** "Progress in the Ruminant Telomere-to-Telomere (RT2T) project"
9:10 p.m. – 9:30 p.m.	**Bekim Sadikovic, London Health Sciences Centre Research Inc** "Novel Integrated Platform for Simultaneous Detection of Genetic Variants and Epigenetic Signatures in Rare Undiagnosed Genetic Diseases"
Wednesday, February 26	**(Morning)**
7:30 a.m. – 9:00 a.m.	**Breakfast Sponsored by Roche** Sunset Terrace
7:30 a.m. – 6:00 p.m.	**Hospitality Desk** Calusa Foyer
Plenary Session:	**Evolutionary Genomics (Beth Shapiro, University California, Santa Cruz, Session Chair)** Calusa Ballroom 1-7
9:00 a.m. – 9:30 a.m.	**David Kingsley, Stanford University** "Sticklebacks and host-pathogen evolution"
9:30 a.m. – 10:00 a.m.	**Ed Green, University of California, Santa Cruz** "DNA-based forensics in the genomics era"
10:00 a.m. – 10:30 a.m.	**Jenny Tung, Max Planck Institute for Evolutionary Anthropology** "The social genome: new insights from long-term field studies of wild mammals"

10:30 a.m. – 11:10 a.m.	**Coffee Break Sponsored by Moleculent** Sponsor Promenade – Osprey Foyer
11:15 a.m. – 11:35 a.m.	**James Banal, (Abstract Selected Talk) Cache DNA** "Breaking the ice: Ambient long-term preservation of ultra-long DNA"
11:35 a.m. – 11:55 a.m.	**Jeremiah Minich, (Abstract Selected Talk) Salk Institute for Biological Studies** "Long-read metagenomics enables a new approach to microbiome research (meta-pangenomics) and uncovers novel bacterial strain genetic associations with pediatric undernutrition"
12:15 p.m. – 1:15 p.m.	**Silver 3 Sponsor Workshop & Lunch - Bruker** **Mark R. Munch, PhD., President of Bruker Nano Group, Jude Dunne, PhD., Vice President, New Product Development, Assays and Assay Systems, Bruker Spatial Genomics, Joseph Beechem, PhD., Chief Scientific Officer, Bruker Spatial Biology, Oliver Braubach, PhD., Director of Research & Development, Bruker Spatial Biology, Parambir S. Dulai, MD, Director of Precision Medicine, Division of Gastroenterology and Hepatology, Northwestern University, Todd Garland, President, Bruker Spatial Biology** "Expanding the limits of spatial biology: Comprehensive integrated technologies for spatial multi-omics" Calusa Ballroom 1-7
12:15 p.m. – 1:30 p.m.	**AGBT Lunch** Sunset Terrace, Lanai Beach
Wednesday, February 26	**(Afternoon)**
Plenary Session:	**Technology (Stacey Gabriel, Broad Institute of MIT and Harvard, Session Chair** Calusa Ballroom 1-7

1:30 p.m. – 2:00 p.m.	**A conversation with John Shoffner, Astronaut, and Chris Mason, Weill Cornell Medical Center**
2:00 p.m. – 2:30 p.m.	**Joe DeRisi, University of California, San Francisco** "Genomic approaches: from infectious to autoimmune disease"
2:30 p.m. – 3:00 p.m.	**Andrew Pask, The University of Melbourne** “Marsupial genomics; paving the way for conservation, preservation and restoration of lost biodiversity”
3:00 p.m. – 3:20 p.m.	**Miranda E. Orr (Abstract Selected Talk), Washington University School of Medicine** “Highest-plex single-cell spatial whole transcriptome and 1000+ proteome identify a new cell type and neural state in Alzheimer’s disease brains”
3:20 p.m. – 3:40 p.m.	**Baoxu Pang, (Abstract Selected Talk) Leiden University Medical Center** “Genome-wide Cas9-mediated screening of essential non-coding regulatory elements via libraries of paired single-guide RNAs”
3:40 p.m. – 3:50 p.m.	**Closing Comments**
7:00 p.m. – Midnight	**Farewell Dinner – 25th Silver Anniversary Sponsored by Watchmaker Genomics** Banyan Ballroom



BRUKER SPATIAL BIOLOGY

Best In Class Solutions For Your
Unique Spatial Biology Needs

Learn
More
Now!

BRUKER

Abstract Table of Contents

Page

MODERNIZE YOUR NGS WORKFLOWS with Watchmaker

WATCHMAKER @ AGBT

NEED A JOLT?

Espresso's on us —
swing by and power up!

Monday & Tuesday, 2 - 4 PM
Osprey 9

PRESENTATION

The Power of Positive:
TAPS Methylation Analysis Unlocks Comprehensive Tumor Multiomics

Tuesday, 2:20 - 2:40 PM

SCIENTIFIC POSTER

TET-assisted pyridine borane sequencing enables the simultaneous detection of cytosine methylation and genetic variants in clinically relevant samples

Scan to learn more or visit watchmakergenomics.com

Retro is chic – but not for library prep!

Upgrade your NGS preps with solutions that are sensitive, rapid, and highly scalable:

- Unlock more information from low-input and degraded samples
- Library prep in 5 hours or less – even for RNA-seq!
- Boost sequence accuracy with fewer artifacts and PCR errors
- Generate PCR-free libraries from as little as 75 ng
- Automation challenges? We have the products – and team – to help!

Visit us in
Osprey 9!

Cancer Omics

Characterizing renewable somatic reference samples developed from clonal engineered cell lines

Adetola Abdulkadir[1], Camille Daniels[1], Nathan D. Olson[2], Justin Zook[2]

1 Medical Device Innovation Consortium (MDIC), Clinical Diagnostics, Arlington, VA
2 National Institute of Standards and Technology (NIST), Material Measurement Laboratory, Gaithersburg, MD

The Somatic Reference Samples (SRS) Initiative, led by the Medical Device Innovation Consortium (MDIC), aims to improve the process for developing and validating next-generation sequencing (NGS)-based cancer diagnostics through the development of reference samples. The lack of well-characterized, validated reference samples has hindered the efficient development of these tests. The SRS Initiative seeks to create a sustainable model for somatic reference sample generation in collaboration with the NIST Genome In A Bottle (GIAB) team. For the SRS pilot project ten clones were engineered to introduce a single clinically relevant cancer variant each into the highly characterized GIAB HG002 cell line via CRISPR (Cas9). Here, we present the development and preliminary characterization of the edited clonal cell lines. Edits ranging from SNVs to gene fusions were confirmed using ddPCR and Sanger sequencing. Whole genome sequencing (WGS) data was generated for the unedited parent cell line (negative controls) along with the ten engineered HG002 clones. RNAseq data was also generated for the cell lines with engineered gene fusions. The sequencing data is being used to confirm on-target edits and detect large unintended copy number changes that may result from the engineering and cloning process. The analysis and characterization of these engineered HG002 cell lines will inform selection of clones for the SRS pilot sample mixture design. Clones selected for inclusion in the mixtures will undergo additional long and short-read sequencing for further characterization. These somatic reference samples will be produced in formalin-fixed, paraffin-embedded (FFPE) format and be subjected to external clinical validation to ensure fitness for use in cancer diagnostics. Use of these samples will aid in NGS-based cancer diagnostic development by facilitating validation and informing assay optimization, and technology development. As a result, their use will expedite regulatory review and therapeutic development.

Integrated multiomic and multimodal analysis of cancer using Todai OncoPanel 2

Hiroyuki Aburatani[1], Katsutoshi Oda[2], Motohiro Kato[3]

1 The University of Tokyo, Genome Science and Medicine Laboratory, Research Center for Advanced Science and Technology, Tokyo, Japan
2 The University of Tokyo, Division of Integrative Genomics, Graduate School of Medicine, Tokyo, Japan
3 The University of Tokyo, Department of Pediatrics, Graduate School of Medicine, Tokyo, Japan

Integrated diagnosis has become increasingly important in determining treatment strategies for individual patients as more treatment options become available. We have developed Todai OncoPanel 2 (TOP 2), a DNA-RNA dual panel assay that analyzes 737 genes for DNA and 455 genes for RNA fusions, as a prototype for the GenMineTOP assay, an approved genetic test by the national health insurance. Additionally, it measures gene expression for approximately 1,400 genes, enabling multiomic analysis by integrating both DNA and RNA data. Transcriptome data is particularly useful for predicting the tissue origin of tumors and identifying tumor subtypes.

Moreover, digital pathology data are collected as H&E images from the tumor specimens used for molecular analysis, minimizing the effects of intratumor heterogeneity and providing a multimodal assay platform. For example, spatial information on tumor-infiltrating lymphocytes can be valuable in predicting the prognosis of MSI-positive patients.

We hope this integrated data platform will offer a robust foundation for developing more precise diagnostic markers and personalized treatment strategies. Data from the ongoing studies will be presented.

Uncovering tumor heterogeneity in NEPC and CRPC using digital spatial transcriptomics profiling: NPTX1 as a diagnostic marker

Alicia Alonso[1,2,3], Paul Zumbo[4,5], Friederike Dundar[4,5], Roberto De Gregorio[1,5], Michael Sigouros[1], David Wilkes[1], Francesca Khani[5], Olivier Elemento[1,5], Doron Betel[3,4,5], Himisha Beltran[7]

1 Weill Cornell Medicine, Caryl and Israel Englander Institute for Precision Medicine, New York, NY
2 Weill Cornell Medicine, Epigenomics Core, New York, NY
3 Weill Cornell Medicine, Department of Medicine, New York, NY
4 Weill Cornell Medicine, Applied Bioinformatics Core, New York, NY
5 Weill Cornell Medicine, Department of Physiology and Biophysics, New York, NY
6 Weill Cornell Medicine, Department of Pathology and Laboratory Medicine, New York, NY
7 Dana-Farber Cancer Institute, Department of Medical Oncology, Boston, MA

NEPC is an emerging variant of advanced prostate cancer, currently diagnosed based on tumor morphology. It shares similar genomic features with CRPC-Adeno, reflecting a luminal prostate cell origin, but is distinguished by specific transcriptional changes. We hypothesize that both metastatic CRPC and NEPC tumors contain phenotypically heterogeneous cellular populations, a complexity not fully captured by immunohistochemistry in pathology or bulk RNA sequencing.

Here, we introduce a proof-of-principle approach and computational framework for utilizing digital spatial profiling to characterize distinct tumor and stromal populations within NEPC and CRPC metastases. This was achieved using the 10x Genomics Spatial Gene Expression workflow for FFPE and for fresh frozen tissues. We determined that the novel NEPC biomarker NPTX1 characterized NEPC cells more precisely than the traditional immunohistochemical markers CHGA and INSM. The spatial transcriptomics workflow could improve the diagnostic accuracy of NEPC cells from biopsied CRPC-Adeno tissues.

Nanopore sequencing of cerebrospinal fluid from individuals with brain metastases reveals a cancer methylation, hydroxymethylation, and fragmentation profile

Tianqi Chen[1], Billy, T. Lau[1], Xiangqi Bai[1], Melanie Hayden Gephart[2], Hanlee P. Ji[1,3]

1 Stanford University School of Medicine, Department of Medicine, Division of Oncology, , Stanford, CA
2 Stanford University, Department of Neurosurgery, Stanford, CA
3 Stanford University, Department of Electrical Engineering, Stanford, CA

As part of the tumor biological process referred to as metastasis, cancers spread from their original site to different organs such as the brain. Individuals with brain metastasis have a particularly poor prognosis denoted by a shortened lifespan. Among the different type of malignancies, lung cancers have the highest risk for brain metastasis. DNA sequencing of cell free DNA (cfDNA) from the blood of cancer patients can detect circulating tumor DNA (ctDNA) from metastatic cancers. However, sequencing cfDNA from blood has limited sensitivity in detecting brain metastasis – an anatomical structure called the blood brain barrier limits the circulation of ctDNA into the blood stream. Cerebrospinal fluid (CSF) surrounds the brain and serves as a better source for enriched ctDNA from brain metastasis.

We studied the genomic properties of cfDNA from the CSF of individuals with lung cancer brain metastases compared to the CSF of healthy controls. Specifically, we used a nanopore-based single molecule sequencing approach to characterize CSF's cfDNA patterns of fragmentation, methylation and hydroxymethylation. This single molecule approach had multiple advantages in determining nucleosome fragmentation patterns with longer sequences and directly deriving methylation genomic features without the use of any chemical or biochemical processing.

In terms of cfDNA fragmentation, mono-nucleosome levels were enriched with a significantly higher mono-/tri-nucleosome ratio for patients with brain metastasis versus controls. We also compared the fragmentation profiles between CSF and plasma to investigate the uniqueness of the fragmentation signature in CSF-derived cfDNA. Comparison with plasma-derived cfDNA confirmed the mono-/tri-nucleosome ratio as a unique fragmentation feature to CSF.

Methylation and hydroxymethylation profiles were obtained simultaneously directly from the CSF cfDNA. Distinct patterns of differential methylation were observed between patients with brain metastases and controls. For hydroxymethylation, we observed significantly lower methylation levels of genome-wide 5-hydroxymethylcytosine CpG (5hmCG) among cancer patients. We identified a set of genes with differential 5hmCG levels that differentiated between CSF from cancer patients and controls. Overall, this proof-of-concept study demonstrated that genomic profiles fragmentation, methylation and hydroxymethylation features of CSF-derived cfDNA provides a biomarker for brain metastasis in lung cancer.

A trio-based Genome in a Bottle mosaic benchmark using multiple sequencing technologies

Camille Daniels[1], Adetola Abdulkadir[1], Nathan D. Olson[2], Justin Zook[2]

1 Medical Device Innovation Consortium (MDIC), Clinical Diagnostics, Arlington, VA
2 National Institute of Standards and Technology (NIST), Material Measurement Laboratory, Gaithersburg, MD

Mosaic and somatic mutations can cause disease but need benchmarks that enable evaluation of low frequency variant calling methods. To address this need, we generated the first mosaic variant Genome in a Bottle benchmark set for HG002. We first used high coverage whole genome sequencing data for the Ashkenazi Jewish (AJ) trio whole genome sequencing (WGS) data (son - HG002, mother - HG003, father - HG004) to identify low frequency variants in the highly characterized HG002 (son) lymphoblastoid cell line. A limit of detection (LOD) of 5% VAF was established using in-silico mixtures of HG002 and HG003 high-coverage data. To generate candidate mosaic variants, somatic variant calling was performed with high coverage (300X) Illumina data (son as tumor and combined parents as normal). After excluding variants in the GIABv4.2.1 germline small variant benchmark, we applied a series of heuristics to filter 366,728 potential mosaics and performed manual curation using the AJ trio along with orthogonal HG002 short- and long-read datasets. 85 variants and 2.45 Gbp of GRCh38 were included in the v1.1 benchmark, where variants were required to fall between 5% and 30% VAF, not have evidence of sequencing or mapping bias, and be absent from the parents.

External evaluation of the draft HG002 mosaic benchmark by eight mosaic or somatic variant calling groups informed refinement, ensuring the benchmark adhered to the reliable identification of errors (RIDE) principle across a variety of variant callers and sequencing platforms. This benchmark set can be used to evaluate mosaic variant callers, as a negative control in WGS somatic calling or targeted clinical sequencing in tumor only mode, benchmark somatic calling in tumor-normal mode using GIAB mixtures, remove mosaic variants from the HG002 germline benchmark, or evaluate off-target edit detection methods. The HG002 mosaic benchmark v1.1 is now available at the GIAB ftp site: https://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/release/AshkenazimTrio/HG002_NA24385_son/mosaic_v1.10/GRCh38/SNV/.

Unraveling TKI resistance in NSCLC cells via RNA, protein, and morphological responses with AVITI24

Vivian Dien[1], Connor Thompson[1], Daniel Honigfort[1], Tristin Rammel[1], Grace Yeo[1], Marielle Krivit[1], Ramreddy Tippana[1], Jane Yao[1], Matthew Kellinger[1], Amirali Kia[1], Sinan Arslan[1], Michael Previte[1]

1 Element Biosciences, Inc., Research and Development, San Diego, CA

Non-small cell lung cancer (NSCLC) is one of the most prevalent and deadly types of lung cancer, with epidermal growth factor receptor (EGFR) mutations being a common driver of tumorigenesis. Targeted therapies such as Gefitinib, a first-generation EGFR tyrosine kinase inhibitor (TKI) have demonstrated clinical efficacy. However, drug resistance and tumor heterogeneity remain critical challenges for fully targeted therapy. A common resistance mechanism is the EGFR T790M mutation, which enables ATP to out-compete first-generation TKIs. Next-generation drugs such as Osimertinib covalently bind to the domain and have been shown to be more effective and selective towards first generation TKI resistant tumors. To investigate the molecular and morphological changes in NSCLC cells in response to TKI therapy, we leveraged the AVITI24 cytoprofiling omics platform, for high resolution, multiplexed analysis of RNA, protein, and cellular architecture. Using AVITI24 cytoprofiling, we characterized three NSCLC lines treated with either Gefitinib or Osimertinib with the resolution to filter at a single-cell level. Within the heterogeneous population, we identified distinct RNA and protein expression patterns between cells treated with either drug. When comparing either Gefitinib-resistant or sensitive cells, distinct levels of EGFR, ERK, and other targets within the MAPK pathway indicated variable inhibition of cell proliferation. A549 cells treated with Gefitinib exhibited up to a Log2FC of -0.108 for phosY1068-EGFR and -0.237 for phosTY-ERK12, indicating significant inhibition, whereas NCI-H1299 and NCI-H1975 on average showed notably less change. Distinct morphological changes were also detected, suggesting drug-specific effects on cell architecture. We then leveraged the ability to convert a highly heterogenous system into its simplest form—a single cell— to increase sensitivity to responses that were not detected within the well-level bulk population. Our findings demonstrate the power of cytoprofiling technologies to elucidate the complex molecular and morphological responses of NSCLC cells to targeted therapies at a granular level, furthering our understanding of the on- and off-target impacts of cancer therapies and the mechanisms of action for Gefitinib and Osimertinib. These findings lay the groundwork for future studies aimed at improving personalized treatment strategies in NSCLC and overcoming the obstacles posed by drug resistance.

Gene fusion and variant-aware isoform detection with functional prediction from long-read RNA sequencing in osteosarcoma

Colette Felton[1], Andrea Galvez[1], Leanne Sayles[1], Christopher Vollmers[1], Alejandro Sweet-Cordero[2,3], Angela N. Brooks[1]

1 University of California Santa Cruz, Department of Biomolecular Engineering, Santa Cruz, CA
2 University of California San Francisco, Department of Pediatrics, San Francisco, CA
3 University of California San Francisco (UCSF), Helen Diller Family Comprehensive Cancer Center, San Francisco, CA

Although most cancer variant profiling is done with short-read-based methods, many cancers are driven by structural variants that are difficult to detect with these methods. Our current approach to understanding driver mutations is also limited to single variants and rarely considers the context in which they are expressed. Alterations in the expression and splicing of genes containing variants can both impact their tumorigenicity and allow them to develop resistance to therapies. We present FLAIR3, which generates a custom transcriptome from long-read RNA sequencing and identifies SNVs, insertions, deletions, and gene fusions from alignment to this transcriptome. We show that this improves accuracy above alignment to the genome or annotated transcripts. FLAIR3 then integrates these variants with the transcripts to predict functional changes to the amino acid sequence. We apply this approach to patient-derived osteosarcoma cell lines, a cancer whose pathology is driven by complex structural variation. In these samples, we identify more complex changes to previously identified amplified drivers such as alternative splicing of MYC and alternative splicing of a gene fusion in CCNE1. We also identify a novel set of actionable cancer gene alterations not previously detected by short-read methods. This includes a large deletion in KEAP1 and a number of novel gene fusions, including TP53 and other genes fused with intergenic regions and a complex 4-locus fusion in NOTCH1 not detected by any preexisting tools. This analysis shows that long-read RNA sequencing can detect novel variants in actionable cancer genes and that integrating splicing and variant alterations provides the most complete picture of gene alterations in cancer.

scRNA-seq from FFPE tissue resolves tumor microenvironment of pre-cancerous pancreatic cysts

Ashley Fletcher[1], Erika Crosby[1], Daniel Nussbaum[1], Margaret O'Connor[1], Elishama Kanu[1], Peter Allen[1]

1 Duke University Medical Center, Department of Surgery, Durham, NC

Intraductal papillary mucinous neoplasms (IPMNs) are precancerous cysts representing the only radiographically identifiable precursor to pancreatic ductal adenocarcinoma (PDAC). Successful interception of PDAC requires identifying high-risk disease, which accounts for 15-20% of IPMNs. Currently, this is only achieved by detecting high-grade dysplasia (HGD) in the epithelial lining of the cyst, which is challenging due to the heterogeneous nature of IPMNs; minute foci of HGD can lie directly adjacent to predominantly low-grade dysplasia (LGD), making it nearly impossible to detect HGD without histopathological review of a surgically resected lesion. Additionally, PDAC is often detected only after surgical removal as well. This heterogeneity has hindered efforts to accurately profile pure HGD or invasive regions, resulting in failed attempts to identify diagnostic biomarkers.

The predominant pathology in many IPMN lesions is LGD. Therefore, we hypothesized that if there are differences between the LGD lesions of low-risk, high-risk, and invasive IPMNs, we could discover a diagnostic biomarker without needing to detect HGD or PDAC specifically. While single-cell RNA-sequencing (scRNA-seq) presents an attractive approach for characterizing IPMNs, it has been limited by the inability to sequence samples with confirmed pathology. Recently, 10X Genomics introduced a scRNA-seq method for formalin-fixed, paraffin-embedded (FFPE) tissues, mitigating the challenge of sequencing samples prior to histological annotation. We employed this scRNA-seq assay on 13 FFPE tissue blocks confirmed to harbor LGD epithelium only: five from patients with only LGD within the entire specimen (low-risk), four from patients with focal HGD detected elsewhere in the specimen (high-risk), and four from patients who had progressed to malignant disease (invasive). We also included two samples containing LGD, HGD, and PDAC for comparative analysis.

Overall, we profiled >100,000 cells, the largest IPMN single-cell sequencing effort to date. We identified a cell population indicative of acinar-to-ductal metaplasia (ADM) in low-risk samples that disappeared through disease progression. Moreover, we discovered that populations of immune cells differed between low-risk, high-risk, and invasive IPMNs. For instance, low-risk samples had the highest proportion of iCAFs and CD8+ TEMs. These results demonstrate the first instance of resolving and comparing the microenvironment surrounding LGD lesions in low-risk, high-risk, and invasive IPMNs.

Identification of Putatively Causal DNA Methylation Markers in Prostate Tissue For Prostate Cancer Risk: Comprehensive Analyses Using Novel meQTL Data as Genetic Instruments

Adrian Gomez[1], Shuai Liu[2], Lang Wu[2]

1 The University of Hawaii at Manoa, Molecular Biosciences and Bioengineering, Honolulu, HI, U.S.A.
2 The University of Hawaii Cancer Center, Honolulu, HI, U.S.A.

DNA methylation is well-established to play an important role in the etiology of prostate cancer, a common malignancy in men. On the other hand, it is challenging to identify putatively causal DNA methylation markers in prostate tissue for this common cancer, due to several common biases existing in conventional epidemiological studies. Recent availability of large-scale reference prostate tissue DNA methylation profiles along with genome information provide significant opportunities to uncover such causally relevant markers using advanced statistical learning approaches. In our prior work, we identified specific methylation biomarkers in blood that are potentially related to prostate cancer regulation by using genetic instruments. While such work provides important implications for the role of methylation in prostate cancer etiology, there remains a significant potential to further enhance the etiology understanding of prostate cancer by uncovering prostate tissue-specific methylation markers through the use of normal prostate tissue-specific datasets. In this project, we aim to employ a novel study design combined with advanced statistical genetic approaches to further investigate methylation markers in prostate tissue related to prostate cancer, by utilizing data from 108 healthy male subjects derived from the Genotype-Tissue Expression (GTEx) project. Our objectives include identifying methylation quantitative trait loci (meQTL) in prostate tissue and then identify prostate cancer risk-relevant methylation markers. We will first explore both cis- and trans-acting prostate tissue-specific meQTL using Genome-Wide Association Study (GWAS) analyses. Subsequently, Two-Sample Mendelian Randomization (TSMR) will be employed to identify CpG sites associated with prostate cancer risk using instrumental variables from both cis- and trans-meQTL. This project aims to substantially advance our understanding of the etiology of prostate cancer through novel applications of advanced statistical genetic tools.

Correlation of a novel extracellular vesicle-based ovarian cancer blood test with patient-matched tumor multiomics analysis of high-grade serous ovarian carcinoma

Toumy Guettouche[1], Gabrielle N. Barcaskey[1], Katherine S. Yang[1], Emily S. Winn-Deen[1], Michael DeRan[2], Daniel Salem[1], Sanchari Banerjee[1], Laura T. Bortolin[1], Delaney M. Byrne[1], Lauren T. Cuoco[1], MacKenzie S. King[1], Troy Hawkins[1], Dawn R. Mattoon[1], Jessica N. McAlpine[3], David Huntsman[4], Anthony D. Couvillon[1]

1 Mercy BioAnalytics, Inc., Waltham, MA
2 Diamond Age Data Science, Somerville, MA
3 Vancouver Coastal Health (VCH), University of British Columbia, Vancouver, British Columbia, Canada
4 Vancouver General Hospital, BC Cancer and University of British Columbia, Vancouver, British Columbia, Canada

We previously demonstrated that detecting colocalized membrane biomarkers on single tumor-associated extracellular vesicles (EVs) provides high specificity and sensitivity for ovarian cancer detection. Our blood-based Ovarian Cancer Test (OC Test) captures EVs and detects five cancer biomarkers --Bone Marrow Stromal Antigen 2 (BST-2), Folate Receptor Alpha (FOLR1), Mucin-1 (MUC-1), Mucin-16 (MUC-16), and sialyl-Tn Antigen (sTN)--which are overexpressed in high-grade serous ovarian carcinoma (HGSC), the most common and aggressive form of ovarian cancer.

In this study, plasma samples from 58 HGSC patients (17 Stage I, 30 Stage II, and 10 Stage III) and 17 benign ovarian tumor patients were obtained from the Ovarian Cancer Research Program (OVCARE, Vancouver, BC) and evaluated with the OC Test. In addition, RNA sequencing was performed on patient-matched formalin-fixed paraffin-embedded (FFPE) tissue sections to assess the correlation between OC Test signal and RNA expression. Since sTn is a cancer-associated carbohydrate, the enzyme alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 1 (ST6GALNAC1), responsible for sialylation, was used as a proxy for sTn RNA expression.

A tumor microarray (TMA) was created and stained using an Opal multiplex-IHC assay, and biomarker colocalization was quantified using AI-based image analysis software. RNA sequencing analysis was performed on 50 FFPE tissue samples (16 benign, 28 early-stage, and 6 late-stage cancers) with > 50% tumor content.

Most benign tissues showed low RNA expression of MUC1, MUC16, FOLR1, and ST6GALNAC1, with intermediate BST2 levels, while HGSC tissues displayed elevated expression of these markers, except BST2, which was downregulated in 25% of cases. Two of the 28 early stage HGSC cases had negative OC Test scores and grouped with benign tissues based on RNA expression. mIHC staining of the TMA and Single Molecule Localization Microscopy (SMLM) analysis of blood EVs for the five biomarkers showed biomarker colocalization in tissue and on individual EVs strongly correlated with OC Test results in both benign and HGSC cases.

These results demonstrate that the OC Test, which detects colocalized biomarkers on EVs in blood, correlates with biomarker expression in primary tumors at both RNA and protein levels, reinforcing the viability of EVs as a liquid biopsy analyte class for early cancer detection.

Separation and parallel analysis of circulating DNA and RNA from liquid biopsies

Fanli Meng[1], Caryn R. Hale[1], Grittney Tam[1], Sitharam Ramaswami[1], Juan Blanco Heredia[1], Michael F. Berger[1], Brian Loomis[1]

1 Memorial Sloan Kettering Cancer Center, Marie-Josée and Henry R. Kravis Center for Molecular Oncology, New York, NY

Analysis of cell-free DNA (cfDNA) from non-invasive liquid biopsies has enabled novel methods for cancer detection and monitoring. However, circulating DNA is only one analyte present in plasma; precious information about RNA, proteins or intact cells is typically lost during cfDNA extraction. Cell-free RNA (cfRNA) is a largely untapped resource in liquid biopsies and may be an effective means for detection of highly expressed fusions and/or expression profiles from circulating tumor material. In order to maximize the information from liquid biopsy samples, we developed a protocol to separate and characterize both cfDNA and cfRNA from a single plasma sample. To do this, total nucleic acids are isolated from the plasma using standard methods, followed by immunoprecipitation of cfDNA using an antibody against double-stranded DNA. This cfDNA goes directly into on-bead library prep and capture via a targeted cfDNA panel. The supernatant, now depleted of cfDNA, is DNase treated, leaving relatively pure cfRNA. This is then subject to RNA library prep, followed by capture via either an exome- or fusion-targeted panel. Using pooled patient plasma, we show that this protocol is fast and efficient, and provides data that is equivalent in quality to either cfDNA or cfRNA extraction alone. We are currently testing the utility of this method to sequence and characterize cfDNA and cfRNA in parallel from individual patient samples. Clinical implementation of this protocol will not only increase the sensitivity of fusion detection from liquid biopsy sources, but also allows for multi-omic approaches such as correlating the presence of cfDNA with gene expression profiles from the source tumor material.

Nanopore-based DNA and single-cell multi-omics reveal that the leukemia KMT2A translocation produces aberrant methylation haplotypes with different fusion gene isoforms

Hanlee Ji[1],[2], Raegan Wood[1], Xiangqi Bai[1], Billy T. Lau[1], Susan M. Grimes[1], Ravi Majeti[2], Giwon Shin[1],[4]

1 Stanford University School of Medicine, Department of Medicine, Division of Oncology, Stanford, CA
2 Stanford University, Department of Electrical Engineering, Stanford, CA
3 Stanford University School of Medicine, Department of Medicine, Division of Hematology, Stanford, CA
4 Department of Genetics, Washington University, St. Louis, MO

Translocations of the KMT2A gene are among the most frequent rearrangements in acute leukemias. Located on the chromosome 11q23 locus, KMT2A encodes a H3K4 methyltransferase which plays critical role in the regulation of gene expression in hematopoiesis. This translocation occurs with different partner genes, generates a gene fusion transcript and indicates a high-risk risk leukemia with a poor prognosis. Moreover, the KMT2A gene fusion is an oncogenic driver that is the target of a recently approved therapeutic.

The genomics features of this leukemia translocation are complex, entailing dramatic structural changes across the locus as well as leukemia-specific epigenetic modifications. Conventional short read sequencing approaches do not reveal many of these critical features. To improve the characterization of this translocation, we used an integrated multi-omic approach on the same cancer. It included: (1) nanopore long read DNA sequencing of the translocation; (2) single cell gene expression and chromatin accessibility; (3) single cell long read sequencing of the chimeric fusion product. The nanopore genomic DNA analysis provided both the break points as well as overlapping methylation haplotypes for both 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC). The single cell analysis provided the related chromatin accessibility alterations, gene expression and full transcript characteristics of the KMT2A fusion chimera.

With this approach we analyzed a set of patient acute leukemias. Nanopore DNA sequencing identified the translocation's exact breakpoint with base pair resolution, haplotype contigs spanning the gene fusion locus, somatic Mb haplotypes and cancer-specific 5mC or 5hmC methylation patterns. We used these results to extrapolate that each leukemia had unique methylation haplotypes spanning the translocation locus and specific to the fusion partner gene. From the same cancers, the single-cell multi-omics revealed how the 5-mC or 5-hmC altered chromatin accessibility and gene expression. Single cell nanopore sequencing showed that each individual leukemia expressed different transcript isoforms of the fusion gene. These new results revealed how translocations lead to an unexpectedly diverse patterns of rearrangement methylation haplotypes and fusion isoforms. This added layer of genomic complexity of translocation may contribute to how effective targeted agents impact these leukemias.

Novel algorithmic toolkit for long-read profiling of tumors reveals hidden somatic variation and improves genetic diagnosis of pediatric leukemia patients

Ayse G. Keskus[1], Jimin Park[6], Daniel E. Cook[10], Lisa A. Lansdon[2], Tanveer Ahmad[1], Asher Bryant[1], Byunggil Yoo[2], Sergey Aganezov[3], Anton Goretsky[1,4], Ataberk Donmez[1,4], Isabel Rodriguez[5], Yuelin Liu[1,4], Xiwen Cui[1], Joshua Gardner[6], Brandy McNulty[6], Samuel Sacco[6], Jyoti Shetty[7], Yongmei Zhao[8], Bao Tran[7], Giuseppe Narzisi[9], Adrienne Helland[9], Pi-Chuan Chang[10], Alexey Kolesnikov[10], Andrew Carroll[10], Erin K. Molloy[4], Chengpeng Bi[2], Adam Walter[2], Margaret Gibson[2], Irina Pushel[2], Erin Guest[2], Tomi Pastinen[2], Karen H. Miga[6], Salem Malikic[1], Chi-Ping Day[1], Nicolas Robine[9], Cenk Sahinalp[1], Michael Dean[5], Kishwar Shafin[10], Midhat S. Farooqi[2], Benedict Paten[6], Mikhail Kolmogorov[1]

1 National Institutes of Health (NIH), National Cancer Institute, Center for Cancer Research, Cancer Data Science Laboratory, Bethesda, MD
2 University of Missouri-Kansas City School of Medicine, Children's Mercy Hospital, Kansas City, MO
3 Oxford Nanopore Technologies, New York, NY
4 University of Maryland, Department of Computer Science, College Park, MD
5 National Institutes of Health (NIH), National Cancer Institute, Division of Cancer Epidemiology and Genetics, Rockville, MD
6 University of California Santa Cruz, Genomics Institute, Santa Cruz, CA
7 Frederick National Laboratory for Cancer Research, Sequencing Facility, Cancer Research Technology Program, Frederick, MD
8 Frederick National Laboratory for Cancer Research, Sequencing Facility Bioinformatics Group, Biomedical Informatics and Data Science Directorate, Frederick, MD
9 New York Genome Center, New York, NY
10 Google Inc, Mountain View, CA

Most human cancers arise from somatic alterations in the cancer genome. Scalable whole-genome sequencing of diverse tumor samples has paved the way to identify the pan-cancer landscape of these genomic changes, including small variants, copy number alterations (CNA), and structural variations (SV). However, current large-scale whole-genome sequencing initiatives primarily utilize short-read sequencing technologies. While long-read sequencing offers significant advantages in terms of mappability and direct haplotype phasing, existing long-read methods were not designed for the complexities of tumor genomes, such as complex rearrangements, clonal heterogeneity, and aneuploidy.

(continue to page 39)

(continued from page 38)

Here, we present a comprehensive toolkit designed for the complete analysis of somatic alterations; in the cancer genome, applicable to both tumor/normal pairs and tumor-only studies. This toolkit includes DeepSomatic (https://github.com/google/deepsomatic) for deep learning-based small variant calling, Severus (https://github.com/KolmogorovLab/Severus) for identifying both simple and complex SVs using breakpoint graphs, and Wakhan (https://github.com/KolmogorovLab/Wakhan) for assessing haplotype-specific CNVs and LOH with improved precision compared to short-read counterparts. All tools are open-source and freely available.

To benchmark these tools, we sequenced five tumor/normal cell line pairs using short-read (Illumina) and long-read (ONT and PacBio) platforms. Our analysis, based on a high-confidence set of variant calls supported by multiple technologies, revealed that all three tools achieved the highest F1 scores (0.76-0.89, 0.78-0.92, and 0.94-0.98 for Severus, DeepSomatic indels and SNVs, respectively) compared to short- and long-read methods. Complex SV analysis with Severus revealed recurring SV patterns, i.e., chromoplexy and chromothripsis. Additionally, we demonstrate that short-read SV calling methods systematically miss certain classes of variants and produce more false positives compared to long-read methods. We openly release this data to encourage more computational developments.

To highlight the translational potential of long-read sequencing, we applied our toolkit to seventeen pediatric leukemia cases from historically under-sequenced populations with known and unknown genetic subtypes. We successfully identified driver mutations in all five known cases. In 4 out of 12 undiagnosed cases, we identified diagnostic cryptic SVs, such as an ins(11;10)(q23.3;p12p12) forming a KMT2A::MLLT10 fusion, missed by routine clinical approaches. This underscores the potential of our toolkit to significantly enhance our understanding and treatment of cancer.

Validation of a clinical Blended Genome Exome (cBGE) assay for prostate cancer polygenic and monogenic risk in veterans

Katie Larkin[1], Candace Patterson[1], Michael Gatzen[1], Chris Kachulis[1], Matthew DeFelice[1], Matthew Coole[1], Marissa Gildea[1], Jonna Grimsby[1], Sean Hofherr[1], Morgan E. Danoski[3], Charles A. Brunette[3], Roshan Karunamuni[2], Anna M. Dornisch[2], Tyler Seibert[2], Jason L. Vassy[3], Niall Lennon[1]

1 Broad Clinical Labs, Burlington, MA
2 University of California San Diego, La Jolla, CA
3 VA Boston Healthcare System, Boston, MA

Prostate cancer is the most common non-skin cancer among Veterans and remains a leading cause of cancer-related deaths in that population. Current screening with prostate-specific antigen (PSA) testing, can lead to false positives, unnecessary biopsies, and overtreatment. To address these challenges, a novel clinical trial within the Veterans Affairs (VA) healthcare system, the Prostate Cancer, Genetic Risk, and Equitable Screening Study (ProGRESS), aims to modernize prostate cancer screening through genetic testing, utilizing a precision medicine approach.

To achieve this aim, the study is leveraging the Blended Genome Exome (BGE) for genetic testing. BGE is a sequencing method developed to reduce the complexity and cost of traditional screening methods. It marries the benefits of Whole Genome Sequencing (WGS) for single nucleotide variant genotyping, utilizing imputation, with deep Whole Exome Sequencing (WES) to identify pathogenic mutations in the coding region.

This approach combines an exome library and a whole genome library into a single tube for sequencing. Initially launched in a research setting, BGE yields a 35-45x exome with a 2-3x genome, allowing for low-cost population-scale sequencing. However, this research-oriented approach lacks the exome depth required for clinical applications. We have developed a clinical Blended Genome Exome (cBGE) to address this gap. In the cBGE process, the amount of exome library present in the blend is increased. This brings the exome to >80x mean coverage, in line with the standard for commercially available clinical exome products, while still incorporating whole genome data at 2-3x coverage. This sequencing approach balances cost considerations with analytical performance and can be used for multiple clinical applications.

In the ProGRESS study, Veterans aged 55-70 will undergo genetic testing using the cBGE platform. PRS and Monogenic testing will be performed and used to identify those at high risk for prostate cancer, including aggressive disease, enabling personalized screening recommendations.

We will describe our workflow and the results of our PRS and monogenic validation study for ProGRESS, which was carried out to assess the performance of this clinical-grade assay. Despite limitations in diagnostic applications, we will also demonstrate the validity of this approach for genetic risk estimation.

Joint single-molecule cell-free DNA and cell-free RNA nanopore sequencing for cancer detection and analysis

Billy Lau[1], Tianqi Chen[1], Xiangqi Bai[1], Susan Grimes[1], Hanlee Ji[1,2]

1 Stanford University School of Medicine, Division of Oncology, Stanford CA
2 Stanford University, Department of Electrical Engineering, Stanford CA

Liquid biopsies are emerging as non-invasive tools for cancer detection. Current methods typically rely on single omic approaches, such as analyzing either mutations, methylation, structural rearrangements, or gene expression. Comprehensive non-invasive analysis of cancer requires more than one omic modality. For example, DNA and RNA should be analyzed together to both detect cancer and determine therapeutic options. However, analyzing more than one omic type is an ongoing challenge due to the lack of sufficient plasma for multiple assays. This has significant impacts in early cancer detection, where existing methods accordingly suffer from poor sensitivity at earlier stages of malignancy.

Solving these challenges, we developed a novel liquid biopsy approach that jointly analyzes both cfDNA (cfDNA) and cell-free RNA (cfRNA) from plasma. Innovations to our experimental workflow yielded high-quality cfDNA and cfRNA nanopore sequencing libraries, achieving several million single-molecule PCR-free reads from just 500 microliters of plasma. Through a unique polyadenylation scheme, we sequenced all cfRNA types, including non-polyadenylated species, which are typically undetectable by other methods. These features were previously unachievable without an order of magnitude increase in plasma volumes or extensive PCR.

Our cfDNA and cfRNA analysis framework was applied to a cohort of healthy donors and cancer patients with colorectal cancer (CRC), as well as to cancer patients who are currently being monitored for treatment response and cancer progression. For cfDNA, methylation and fragmentomics revealed differential methylation patterns and cfDNA size variations between patients and controls. Consistent with previous studies, we showed that cancer status was associated with increased mononucleosomes and shorter cfDNA fragments. These insights were uniquely enabled by nanopore sequencing. In cfRNA, we also identified differential transcript abundances that were highly specific for CRC, including RNAs that are not normally polyadenylated and would not be found with previous approaches. Some of these transcripts have not been described as biomarkers and could have future utility for cancer detection.

Overall, our methods highlight the capability of nanopore-based cfDNA and cfRNA sequencing to develop comprehensive multi-omic analytics. This joint framework enables a promising new approach for high-accuracy, non-invasive early cancer detection and patient monitoring.

Conserved spatial subtypes and cellular neighborhoods of cancer-associated fibroblasts revealed by high-plex single-cell spatial multi-omics

Yunhe Liu[1], Jimin Min[1], Yibo Dai[1], Luisa M. Solis Soto[1], Humam Kadara[1], Anirban Maitra[1], Linghua Wang[1], Kyung Serk Cho[1]

1 The University of Texas MD Anderson Cancer Center, Houston, TX

Background: Cancer-associated fibroblasts (CAFs) are a multifaceted cell population essential in shaping the phenotype and evolution of the tumor microenvironment (TME). Distinct phenotypic subtypes of CAFs have been identified through advanced molecular profiling techniques, which have been shown to contribute to tumorigenesis, metastasis, and therapy resistance. However, much remains to be understood about how CAFs spatially interact with cancer cells and immune cells within the TME. The recent development of advanced single-cell spatial transcriptomics platforms has opened unprecedented opportunities for in-depth characterization of CAFs.

Methods: In this study, we conducted a comprehensive spatial transcriptomic analysis of CAFs across eight cancer types, using the CosMx and MERSCOPE platforms to profile over 5.7 million cells from 24 large tissue sections as our discovery cohorts. We developed innovative computational method to spatially define four distinct spatial CAF subtypes that are conserved across these cancer types. Using multiple approaches, we examined their phenotypic differences, specific spatial niche organization, and interactions with neighboring immune cells. Additionally, we investigated the impact of CAF subtypes on immune cell infiltration into the tumor core. To validate our findings, we employed multiple spatial platforms, including Visium, Xenium, COMET, CODEX, and IMC, to ensure reproducibility across different cancer types. Beyond validation, with additional samples and clinical data collected, we did comprehensive association analyses to correlate CAF characteristics with patient clinical information and outcomes.

Results: Our approach enabled an unprecedented examination of the spatial phenotypes of cancer-associated fibroblasts (CAFs) in large, complex tissues at high resolution, revealing insights that had not been previously described. We comprehensively characterized CAFs across solid tumor types and identified four conserved spatial CAF subtypes. These subtypes are defined by distinct organizational patterns, neighboring cell compositions, interaction networks, and transcriptomic features. Moreover, they exhibit significant correlations with the abundance, distribution, and phenotypes of tumor-infiltrating immune cells, as well as associations with tumor histology, stage, and patient survival.

Conclusions: Our study highlights the significant correlations between spatial CAF subtypes, levels of immune cell infiltration, states and distribution, and clinical outcomes, providing insights into the spatial heterogeneity and potential roles of CAFs in tumor progression and immune modulation.

ROBIN - Revolutionizing CNS tumor molecular classification with real-time nanopore sequencing and adaptive sampling

M Loose[1], S Deacon[1], I Cahyani, N Holmes, G Fox, R Munro, S Wibowo, T Murray, H Mason, M Housley, D Martin, A Sharif, A Patel, R Goldspring, S Brandner, F Sahm, S Smith[2], SML Paine[2]

1 University of Nottingham, School of Life Sciences, Nottingham, Nottinghamshire, United Kingdom
2 University of Nottingham, Nottingham University Hospital, Nottingham, Nottinghamshire, United Kingdom

ROBIN (Rapid nanpOre Brain intraoperatIve classificatioN) is a groundbreaking tool that utilizes PromethION nanopore sequencing technology for rapid and comprehensive molecular profiling of CNS tumours. Advances in genomic sequencing have revolutionized the diagnostic pathways for brain tumours, shifting focus to epigenetic signatures for classification. ROBIN capitalizes on this shift, offering real-time, intraoperative methylome classification alongside next-day comprehensive profiling within a single assay. This approach enhances diagnostic precision and aids surgeons in tailoring surgical strategies based on immediate genetic insights, thus balancing the risks and benefits of intervention.

ROBIN integrates adaptive sampling through tools like ReadFish, enabling targeted sequencing of critical genomic regions. This method optimizes sequencing efficiency, increasing coverage of key areas while reducing time and cost. It employs a multi-classifier system—incorporating algorithms such as Sturgeon, CrossNN, and RapidCNS2—ensuring high accuracy in methylation-based tumour classification. Classifier performance has been validated on 50 prospective intraoperative cases, achieving robust tumour classifications within minutes and a diagnostic turnaround time of less than 2 hours, with 90% concordance with final integrated diagnoses.

Beyond methylation profiling, ROBIN's ability to generate long reads allows it to detect single nucleotide variants (SNVs), copy number variants (CNVs), and structural variants (SVs) in real time. This capacity is especially valuable for identifying complex fusions that traditional methods may miss, providing a complete integrated molecular classification within 24 hours. While some additional assays may be needed for pathognomonic mutations, ROBIN significantly reduces the time to actionable diagnostic information.

These methods including ROBIN's real-time analysis, combined with ongoing advancements in algorithms, promise to transform the landscape of brain tumour diagnosis, delivering faster, more precise, and actionable results that can directly impact patient care. In addition, ROBIN can dynamically assign target sequences to rapidly stratify samples for candidate trials as well as explore structural variation throughout the genome.

AmpliconSuite: characterizing extrachromosomal DNA and other focal amplifications in the cancer genome

Jens Luebeck[1], Edwin Huang[1], Ted Liefeld[1], Ivy Tsz-Lo Wong[2], Bhargavi Dameracharla[1], Gino Prasad[1], Forrest Kim[1], Paul Mischel[2], Jill Mesirov[1], Vineet Bafna[1]

1 University of California San Diego, La Jolla, CA
2 Stanford University, Stanford, CA

Oncogene amplification in cancer occurs by multiple genomic mechanisms. Not all mechanisms result in the same outcomes. Extrachromosomal DNA (ecDNA) amplifications and breakage fusion bridge amplifications in particular are associated with worse patient survival even when controlling for cancer type. EcDNA, which are acentric, megabase-scale structures, undergo asymmetric segregation during cell division, and in combination with positive selection for oncogenic elements they carry, can reach hundreds of copies per cell. However, distinguishing different amplification modes using whole-genome sequencing (WGS) data is challenging due to their complex copy number and structural variation profiles.

We present AmpliconSuite, a workflow for focal amplification analysis from WGS data, along with AmpliconRepository.org, a companion platform for sharing the workflow's results. AmpliconSuite's core module, AmpliconArchitect, jointly analyzes structural variants and copy numbers, providing insights into ecDNA's role in cancer across thousands of samples. AmpliconSuite streamlines the process by standardizing candidate region identification and deploying AmpliconClassifier for amplification mechanism reporting. We validated these outputs on a cytogenetic dataset of over 150 cancer cell-line/gene FISH probe combinations. The tabular outputs annotate gene content, copy numbers, and summarize structural complexity.

We have recently embedded AmpliconSuite into the GenePattern web interface, allowing users to analyze data directly through a web browser. To date, AmpliconSuite has been deployed across 20+ large-scale studies, analyzing more than 30,000 tumor samples and over 1.5 petabytes of sequencing data.

To foster collaboration and sharing of focal amplification calls, we released AmpliconRepository.org, a community-editable platform for sharing focal amplification results. It offers public ecDNA predictions on over 4,500 tumor samples from datasets like TCGA, PCAWG, and CCLE.

We recently added methods to our workflow to categorize ecDNA types by matching structural variation patterns to formation mechanisms, distinguishing ecDNA generated by chromothripsis from those arising through excision or stalled replication forks. We validated these predictions with data from engineered systems of ecDNA formation.

AmpliconSuite simplifies the identification of focal amplifications and enables public data sharing, while uncovering new insights into the biology of ecDNA and the other unique focal amplification modes, establishing itself as a core tool for oncogene amplification research.

TET-assisted pyridine borane sequencing enables the simultaneous detection of cytosine methylation and genetic variants in clinically relevant samples

Craig Marshall[1], Travis Sanders[1], Aaron Garnett[1], Martin Ranik[1], Josh Haimes[1], Giulia Corbet[1], Kristina Giorda[1], Eduard Casas[1], Thomas Harrison[1], Doug Wendel[1], Ross Wadsworth[2], Eric van der Walt[2], Brian Kudlow[1]

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

Multiomic approaches are gaining recognition for their potential to revolutionize cancer detection and diagnosis by integrating diverse biological data types, including genomic, epigenomic, transcriptomic, proteomic, and fragmentomic modalities. Recent studies highlight the efficacy of multiomics in enhancing the sensitivity and specificity of early cancer detection through liquid biopsies. In particular, DNA methylation, assessed in combination with somatic variants on a whole genome level, is emerging as a powerful predictor of cancer signature origin in cfDNA.

TET-assisted pyridine borane sequencing (TAPS) is a novel, positive readout, method for DNA methylation detection, which uses successive oxidation, reduction, and amplification reactions to convert 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) to thymine. TAPS is a non-destructive chemistry, which enables the simultaneous detection of cytosine methylation as well as genetic variants, while preserving DNA fragment integrity and library complexity. Due to the relative scarcity of CpG dinucleotides in the human genome, the vast majority of the DNA sequence in TAPS libraries is unchanged by the chemistry, which improves sequence quality and greatly simplifies variant detection. This also means that TAPS-treated libraries can be used in target enrichment workflows with standard hybrid capture panels, unlike bisulfite and EM-Seq, which require custom designs.

We have implemented a number of protocol and formulation improvements, resulting in a streamlined TAPS chemistry which delivers a high 5mC conversion efficiency, library complexity retention, and balanced whole genome coverage. In this study, we demonstrate the utility of TAPS in a multimodal setting by simultaneously detecting epigenetic signals and genetic variation in a variety of cancer samples. TAPS allowed the detection of tissue and tumor-specific differentially methylated regions, as well as somatic variants such as C to T mutations, not easily detectable using existing base-level methylation detection technologies. Additionally, we demonstrate the combination of TAPS and target enrichment, to achieve deep and uniform coverage of somatic variants and differentially methylated regions with a single panel. These findings underscore the potential of TAPS as a transformative tool for advancing multiomic cancer detection, enabling precise and comprehensive analysis of both genetic and epigenetic alterations in a single workflow.

Transitioning to clinical whole genome sequencing for germline diagnostics: enhancing assay performance, reporting, and patient care

Karen L. Mungall[1], Xuanjin Cheng[1], David Mulder[1], Eric Chuah[1], Richard Moore[1], Andrew J. Mungall[1], Lucas Swanson[1], Angela Tam[1], Stephen Yip[2], Steven J.M. Jones[1]

1 Canada's Michael Smith Genome Sciences Centre, Vancouver, British Columbia, Canada
2 University of British Columbia, Department of Pathology and Laboratory Medicine, Vancouver, British Columbia, Canada

To improve clinical germline diagnostic services in British Columbia, we are transitioning our hybrid capture Hereditary Cancer Program panel assay to a clinical whole genome sequencing (cWGS) assay utilizing the Illumina DRAGEN platform. This strategy removes probe set design, purchase, and iterative design barriers, whilst also reducing biases such as GC content and amplification artifacts often seen in targeted sequencing. Here we describe a clinically accredited cWGS assay that provides recall of 99.5% for small variants and enhanced CNV detection down to 1kb and indel detection up to 49nt using a minimum de-duplicated coverage depth of 35 fold within a defined virtual panel space.

A challenge of cWGS assays is the vast expansion of sequenced genomic space and the lower coverage which results in technical artifacts and false calls distinct from those observed in hybrid capture. Consequently, we discuss strategies to mitigate these technical challenges and thereby achieve precision of >97%.

The shift to cWGS enables a more comprehensive genomic feature set in the context of a complete genome, but presents interpretation challenges for those used to reporting on higher depth targeted sequencing assays. Thus, we compared the performance of the hybrid capture and cWGS assays in the same target space with the same sample set to identify, evaluate, and mitigate reporting challenges.

A cWGS approach for germline clinical diagnostic testing improves and streamlines processes in the laboratory, quality for interpreters, and care for the patient. We present a cWGS platform that leverages existing sequence data for validation testing to enable the rapid expansion of genomic features and virtual gene target panels needed for population level germline diagnostic testing.

5-methylcytosine and 5-hydroxymethylcytosine are additive biomarkers for early detection of colorectal cancer

Robert Osborne[1], Fabio Puddu[1], Annelie Johansson[1], Aurélie Modat[1], Jamie Scotcher[1], Ricca Sethi[1], Shirong Yu[1,2], Nick Harding[1], Mark Hill[1], Ermira Lleshi[1], Casper Lumby[1], Jean Teyssandier[1], Michael Wilson[1,3], Robert Crawford[1], Tom Charlesworth[1], Shankar Balasubramanian[1,4,5], Páidí Creed[1]

1 biomodal Ltd, The Trinity Building, Chesterford Research Park, Cambridge, United Kingdom
2 Tagomics Ltd, The Cori Building, Little Abington, Cambridge, United Kingdom
3 Princeton University, Princeton, NJ
4 University of Cambridge, Cancer Research UK Cambridge Institute, Cambridge, United Kingdom
5 University of Cambridge, Yusuf Hamied Department of Chemistry, Cambridge, United Kingdom

Early cancer detection has the potential to significantly improve treatment outcomes and survival rates. This study investigates the roles of 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) as biomarkers for early-stage colorectal cancer (CRC) detection in cell-free DNA (cfDNA). Using whole genome sequencing, we analyzed cfDNA from 37 treatment-naive CRC patients and 32 healthy controls. Our findings indicate that combining measurements of 5mC and 5hmC significantly enhances diagnostic accuracy (AUC = 0.95) compared to traditional approaches that conflate these markers (modC). Notably, 71.7% of differentially methylated regions (DMRs) exhibiting an increase in 5hmC in stage I cfDNA also showed a corresponding decrease in 5mC in stage IV, suggesting that 5hmC can effectively track regions of demethylation during tumor development. These results support the hypothesis that distinguishing between 5mC and 5hmC can improve the sensitivity of liquid biopsy tests for early cancer detection.

Targeted NanoSeq: genome-wide detection of somatic mutations at single-molecule resolution

Michael Quail[1], Stefanie V. Lensing[1,2], Federico Abascal[2], Andrew R.J. Lawson[2], Pantelis A. Nicola[2], Iñigo Martincorena[2]

1 Wellcome Sanger Institute, Sequencing Operations, Hinxton, United Kingdom
2 Wellcome Sanger Institute, Cancer, Ageing and Somatic Mutation Programme, Hinxton, United Kingdom

As somatic mutations are only present in a small number of cells, specialised methods with extremely low error rates are required to study them. We previously presented NanoSeq (Nature: https://doi.org/10.1038/s41586-021-03477-4), a duplex-sequencing method with an unprecedentedly low error rate of < 5x10-9 errors per bp in single molecules of DNA. As this error rate is two orders of magnitude lower than somatic mutation rates in human tissues (~10-7), NanoSeq enables the detection of somatic mutations in any tissue type. The use of restriction enzymes constrained NanoSeq to the study of 30% of the human genome, thus limiting driver discovery. We now present a new version of NanoSeq that provides whole genome coverage, whilst maintaining the ultra-low error rate of NanoSeq. This is achieved by combining sonication for genome fragmentation with error-free mung bean nuclease mediated end-repair. This protocol can be combined with bait capture to enrich for regions of interest, enabling whole-exome and targeted NanoSeq applications. As Targeted NanoSeq does not copy errors between strands, this method can also be applied to more damaged DNA samples. We were able to determine a comparable mutational load in highly damaged FFPE samples and high-quality snap frozen samples. In contrast, using traditional duplex-sequencing methods which utilise polymerases, error rates were 10X higher in FFPE tissues compared to snap frozen tissue. Targeted NanoSeq enables the accurate quantification of somatic mutation rates, signatures and cancer-driver mutations in polyclonal tissues with single-molecule sensitivity, giving insights into the earliest stages of disease development, when mutant clones are still microscopic in nature. Thus, Targeted NanoSeq can reveal the polyclonal landscape of any tissue, enabling the study of initiating events in carcinogenesis, ageing and disease. A manuscript describing this methodology and its application at epidemiological scale has been submitted to Nature.

Integrated analysis of cancer whole genome sequencing and whole metagenome sequencing of colorectal cancer

Tatsuhiro Shibata[1],[2]

1 The University of Tokyo, The Institute of Medical Science, Laboratory of Molecular Medicine, Tokyo, Japan
2 National Cancer Center, Division of Cancer Genomics, Tokyo Japan

Background The global incidence of colorectal cancer (CRC) has shown a worrying upward trend, particularly in Japan and certain other regions. The rise in incidence is commonly attributed to an increasingly westernized lifestyle, but the prevalence in Japan exceeds that in the US and Europe, suggesting the presence of other contributing factors. Furthermore, the incidence of early-onset CRC, diagnosed before the age of 50, has been steadily increasing worldwide. A crucial aspect of CRC is its direct connection to the gut microbiota.

Results A total of 138 treatment-naïve Japanese patients diagnosed with histologically confirmed CRC were enrolled. Tumor tissue was subjected to whole-genome sequencing, ribosome-free RNA sequencing and stool samples from the same patient underwent shotgun metagenomic sequencing. Using explanatory artificial intelligence analysis, the patients were classified into four distinct subtypes (onco-microbiome subtypes 1-4). F. nucleatum and P. stomatis significantly influenced subtype 2. In contrast, subtype 1 was characterized by the absence of specific bacterial species. P. stomatis and F. nucleatum were found to be highly distinctive bacterial species in subtype 3 and subtype 4, respectively. This microbiome-based subclassification correlated with somatic genetic events such as hyper-mutation, gene amplification and driver mutations (ARID1A and PTEN). Subtype 3 in particular had a higher proportion of younger patients with an unfavorable prognosis.

Real-time clinical epitranscriptomics improves leukemia diagnosis and unveils long non-coding RNAs as novel therapeutic targets

Martin A. Smith[1,2,3,4], Qing Wang[1], Mélanie Sagniez[3,4], Anshul Budhraja[3,4], Bastien Paré[4], Claire Fuchs[4], Shawn M. Simpson[4], Maxime Caron[4], Elizabeth Snell[5], Etienne Rainmondeau[5], Vincent-Philippe Lavallée[4,6], Thai Hoa Tran[4,6], Alain Bataille[4], Daniel Sinnet[3,4]

1 UNSW Sydney, School of Biotechnology and Biomolecular Sciences, Sydney, New South Wales, Australia
2 UNSW Sydney, RNA Institute, Sydney, New South Wales, Australia
3 Universiy of Montreal, Department of Biochemistry and Molecular Medicine, Montreal, Quebec, Canada
4 Centre Hospitalier Universitaire (CHU) Sainte-Justine Research Centre, Montreal, Quebec, Canada
5 Oxford Nanopore Technologies, Oxford, United Kingdom
6 Universiy of Montreal, Department of Pediatrics, Montreal, Quebec, Canada

Gene expression profiling (GEP) has emerged as a powerful tool for classifying disease subtypes through molecular signatures. Leveraging GEP data from over 1100 pediatric leukemia samples generated via short-read sequencing, we trained a neural network for rapid and accurate disease subtype classification. Applying this model to real-time nanopore sequencing data enables precise disease subtype identification in under 5 minutes, with a cost below $300.

Our subsequent analysis of nanopore cDNA and native RNA transcriptomes uncovered multiple long non-coding RNAs (lncRNAs) among the key diagnostic biomarkers, revealing significant gaps in existing annotations. Further, native RNA sequencing led to the discovery of over 1000 novel lncRNA genes and isoforms containing retro(trans)posons, many overlapping pertinent GWAS loci, suggesting their involvement in disease aetiology.

Using CRISPR interference (CRISPRi) and Cas13-mediated knock-downs, followed by deep differential GEP using RNA004 PromethION flowcells, we validated the oncogenic functions of these lncRNAs, demonstrating their critical functions in gene expression regulation and signaling pathways. Employing the latest Dorado base calling software, we also performed RNA modification analysis for m6A, pseudouridine, inosine and m5C, which identified thousands of isoforms with differential post-transcriptional modifications that did not display significant changes in abundance.

This multi-omic approach integrates differential RNA isoform expression and RNA modifications, offering a robust resource for novel biomarker discovery and personalized therapeutic targeting in leukemia. Our findings highlight the potential of real-time clinical epitranscriptomics to advance cancer diagnostics and treatment strategies.

Comprehensive multi-omics characterization of high-grade neuroendocrine carcinoma of the lung using long-read sequencing and spatial omics techniques

Ayako Suzuki[1], Yoshiki Otsuka[1], Megumi Tateishi[1], Ryota Matsuoka[2], Daisuke Matsubara[2], Yutaka Suzuki[1]

1 The University of Tokyo, Graduate School of Frontier Sciences, Chiba, Japan
2 The University of Tsukuba, Department of Diagnostic Pathology, Faculty of Medicine, Ibaraki, Japan

High-grade neuroendocrine carcinoma (HGNEC) of the lung, including small cell lung cancer and large cell neuroendocrine carcinoma (LCNEC), generally shows poor prognosis. To understand inter- and intra-tumor heterogeneity of molecular features associated with therapeutic difficulties of HGNEC, we performed multilayered analyses for 34 HGNEC cases using long-read sequencing and spatial omics techniques including Xenium.

We first conducted long-read whole-genome sequencing analysis for 21 LCNEC cases and identified various complicated genomic alterations and epigenomic aberrations. We found that large-scale aberrant structures, such as chromothripsis-like events and extrachromosomal DNA candidates, were observed in ASCL1+ neuroendocrine (NE) or POU2F3+ non-NE tumors while YAP1+ non-NE tumors harbored much other types of SVs. At the side of DNA methylation statuses, genome-wide hypomethylation occurred in most of the cases. Especially, DNA methylation statuses in non-NE types were more diverse among the cases.

We further conducted Xenium analysis using the 302-gene lung panel to understand HGNEC transcriptomic heterogeneity. We obtained Xenium data of more than 14 million cells, including NE and non-NE type tumors and their surrounding microenvironmental cells. Several cases harbored both NE and non-NE areas in the tissue section. Based on the results of clustering and co-expression network analysis, tumor cells were mainly classified into three groups showing high YAP1, ASCL1 or ASCL2 expressions, respectively. Interestingly, several tumor sub-clusters showed both NE and non-NE marker expressions. These bivalent signatures were branched out from various states according to trajectory analysis. We are now expanding the analysis to Xenium 5k and Visium HD for more comprehensive elucidation of transcriptome networks regulating HGNEC.

By integrating the complicated genomic/epigenomic aberrations and the network-level transcriptomic features in HGNEC cells, we would precisely understand heterogeneous molecular features that might be associated with therapeutic resistance.

Unveiling genomic heterogeneity and tumor evolution with single-cell DNA sequencing

Josh N. Vo[1], Pankaj Vats[2], Dan R. Robinson[1], Arul M. Chinnaiyan[1],Tong Zhu[2], Hila Benjamin[3], Sarah Pollock[3], Nika Iremadze[3], David Bogumil[3], Ilya Soifer[3], Justin Zook[4]

1 University of Michigan, Department of Pathology, Ann Arbor, MI
2 NVIDIA Corporation, Santa Clara, CA
3 Ultima Genomics, Fremont, CA
4 National Institute of Standards and Technology (NIST), Material Measurement Laboratory, Gaithersburg, MD

Single-cell DNA sequencing (scDNA-seq) has emerged as a transformative technology for elucidating the complexities of tumor evolution and clonal diversity. The Genome in a Bottle (GIAB) consortium recently released extensive bulk WGS and karyotyping data for the new pancreatic ductal adenocarcinoma cell line HG008, demonstrating genomic instability and polyclonality. In this study, we leveraged the Bioskryb single-cell whole genome sequencing (WGS)—Ultima platform to further investigate this heterogeneity at the single-cell level and enhance our understanding of tumor dynamics in HG008.

To achieve this objective, we developed LaScala, a novel computational method specifically designed for analyzing copy number variation (CNV). LaScala integrates log-coverage ratio and B-allele frequency to jointly infer ploidy and purity, demonstrating high sensitivity and specificity in segmentation, even in challenging low-coverage data. We applied LaScala to analyze whole-genome sequencing (WGS) data from 119 individual HG008 cells generated by the Bioskryb ResolveDNA kit and sequenced on the UG100 instrument with coverage ranging from 11 to 45x. Our findings revealed a high level of concordance between bulk and single-cell results concerning gross chromosomal gains, deletions, and losses of heterozygosity. Several subclonal copy number alteration events were identified involving chromosomes 1q, 2, 6p, 10, 15, and X. Additionally, there is evidence of whole genome duplication in more than 5% of the cell population.

We further clustered the single-cell data and identified hierarchical relationships among cell populations within the dataset. Other genomic profiles, including mutations and structural rearrangements, exhibited a strong correlation between bulk tissue and single-cell sequencing, although more subclonal alterations were observed in single-cell data. Additionally, we attempted to reconstruct the phylogenetic relationships among tumor subclones by statistical inference. Overall, our comprehensive approach allowed us to map out the evolutionary trajectories of different clones, offering insights into how intraclonal heterogeneity developed under minimal selection pressure. This work highlights the potential of integrative bulk and single-cell sequencing technologies to advance our understanding of tumor molecular genetics.

Enabling long-read mRNA-seq for biomarker discovery using limited clinical sample inputs

Yue Yun[1], Lisa Welter[1], Kazuo Tori[1], Jackson Peterson[1], Alan Du[1], Gilma Sevilla[1], Yana Ryan[1], Ploy Setthasap[1], Sherry Wei[1], Tomoya Uchiyama[1], Shiyi Yin[1], Mike Covington[1], Mohammad Fallahi[1], Saloni Pasta[1], Bryan Bell[1], Andrew Farmer[1]

1Takara Bio USA, Inc., San Jose, CA

RNAs serve a central role in cellular biology by converting genomic information into effector molecules, either as mRNAs or directly functional RNAs. Measurement of RNA through RNA Sequencing has become an important method to understand biological processes. While it has been long appreciated that a vast diversity of RNA transcripts can be produced through alternative splicing, more recent work has defined a contribution of alternate and aberrant splicing to diseases like cancer, neurodegeneration, and autoimmunity. Thus, understanding the diversity of RNA transcripts is a key emerging topic in pathophysiology and biomarker discovery.

Third-generation sequencing technologies provide the opportunity to sequence full-length cDNA without the need for fragmentation and hence provide a more complete picture of isoform structure and transcript abundance. However, a current limitation of long-read RNA sequencing (LR-RNA-seq) is the requirement for high input amounts of RNA that can be unachievable for some primary sample types, like RNA isolated from small amounts of tumor samples or sorted blood cancer cells.

Here we describe a new LR-RNA-seq product enabling full-length RNA sequencing from single cells (~10pg) up to 100 ng total RNA. With a 10ng total RNA input we demonstrate the ability to reliably sequence at an average length (N50) of 2kb and detection of full-length transcripts as long as 10kb. Performance at the single-cell level substantially outperforms existing single-cell LR-RNA-seq methods. Further, our data provides a more complete picture of isoform-specific changes compared to other commercially available technologies. With the ability to process up to 96 samples at a time, this technology will enable the processing of rare or valuable samples to uncover novel biomarkers beyond gene expression.

Decoding immune networks of the AML bone marrow microenvironment with spatial multi-omic profiling

Katie Maurer[1,2,3], Florian Raths[4], Kenneth H. Gouin III[4], Cameron Y. Park[5,6], Michael Lawson[4], Jake Koh[4], Martin M. Fabani[4], Kenneth J. Livak[1], Jerome Ritz[1], Robert J. Soiffer[1], Elham Azizi[5,6], Eli Glezer[4], Catherine J. Wu[1,2,3]

1 Dana-Farber Cancer Institute, Department of Medical Oncology, Boston, MA
2 Harvard Medical School, Boston, MA
3 Broad Institute of MIT and Harvard, Cambridge, MA
4 Singular Genomics, San Diego, CA
5 Columbia University, Department of Biomedical Engineering, New York, NY
6 Columbia University, Irving Institute for Cancer Dynamics, New York, NY

Spatial profiling of bone marrow may provide unique insights into mechanisms of relapse and maintenance of remission in acute myeloid leukemia (AML) patients after hematopoietic stem cell transplant (HSCT). Interrogating gene expression in bone marrow is challenging due to RNA degradation from decalcification and formalin fixation and paraffin embedding (FFPE). Additionally, the limited throughput of existing platforms is a barrier to profiling clinical-scale cohorts.

Singular Genomics is developing the G4XTM, an in-situ spatial sequencing platform that enables simultaneous profiling of transcriptomics, proteomics, and H&E staining in the same tissue section, over a large imaging area (up to 40 cm2 per run). The methods are designed to work with standard archived samples (FFPE) and are tolerant to fragmented RNA. We used a pre-production G4X to analyze 24 bone marrow biopsies of 12 post-HSCT relapsed AML patients before and after treatment with donor lymphocyte infusions (DLI, 6 responder [R], 6 non-responder [NR]).

Using a targeted panel of 153 immune-related transcripts and 8 proteins, we profiled 193,651 individual cells (median of 30 transcripts, 11 unique genes per cell), revealing the spatial immune cellular networks in the marrow microenvironment. Dimensional reduction and clustering identified 14 cell states annotated using signatures derived from matched single-cell RNA-sequencing data (Maurer et al ASH 2023).

Our previous findings (Maurer et al, ASH 2023) pinpointed cytotoxic CD8+ T cells as a nexus for intra-marrow cellular interactions. In the in situ data, these CD8+ T cells were found near myelo-erythroid progenitor and AML cells, suggesting active immune surveillance post-DLI (NR post: 1.62% vs R post: 6.32%, 32.5 µm radius, p=0.008). This population exhibited decreased spatial self-association in NRs (pre to post: 8.79% to 2.17%, p=0.008). NR marrow pre-DLI was marked by an overabundance of myeloid cells. The microenvironment of Rs showed higher diversity post-DLI (Shannon Index NR post: 0.744 vs. R post: 0.943, p=0.02).

We present the first analysis of post-HSCT relapsed AML marrow spatial multi-omic microenvironment profiling using G4X and find higher immunologic diversity among Rs post-DLI, centering around CD8+ CTLs. Further evaluation of spatial organization may lead to novel targets for optimizing adoptive cellular immunotherapy.

Uncovering a rare BRAF p.V600E mutation in breast cancer through whole-genome sequencing: a precision medicine approach

Youngseok Ju[1], Ryul Kim[1], Ji-Yeon Kim[2], Erin Connolly-Strong[1], Yoon-La Choi[3], Young Seok Ju[1,4], Yeon Hee Park[2]

1 Inocras Inc., San Diego, CA
2 Sungkyunkwan University School of Medicine, Samsung Medical Center, Department of Medicine, Division of Hematology-Oncology, Seoul, Republic of Korea
3 Sungkyunkwan University School of Medicine, Samsung Medical Center, Department of Pathology and Translational Genomics, Seoul, Republic of Korea
4 Graduate School of Medical Science and Engineering, Korea Advanced Institute of Science and Technology, Daejeon, South Korea

Background:
Understanding the genomic landscape is critical for effective cancer management. Targeted panel sequencing, which typically focuses on frequently mutated genes in specific cancer types, may miss rare but clinically significant mutations. Recent advancements and cost reductions in whole-genome sequencing (WGS) provide a comprehensive alternative, allowing for the detection of common and rare mutations, including those in non-coding regions that may influence cancer progression and treatment response. WGS enhances cancer treatment precision by offering a more complete view of the genomic alterations driving disease.

Method:
A 41-year-old female presented with abdominal distension, weight loss, and jaundice. Diagnostic imaging and biopsy confirmed hormone receptor-positive (ER+/PR+/HER2-) breast cancer with metastases to the liver, peritoneum, and lymph nodes. Despite standard treatments, including hormone inhibitors and chemotherapy, her cancer progressed. Initial genetic testing for common mutations associated with breast cancer was negative, complicating treatment decisions. Given the limited response to standard therapies, comprehensive genomic testing using target-enhanced whole-genome sequencing (TE-WGS) (CancerVision, INOCRAS, San Diego, CA) was performed to identify potentially actionable mutations.

Results:
TE-WGS identified a rare BRAF p.V600E mutation, which is typically associated with melanoma and colorectal cancer but is found in only 0.11% of breast cancer cases. This mutation is actionable due to its alignment with BRAF inhibitors, such as vemurafenib, which is FDA-approved for melanoma. Based on this genomic insight, the patient was enrolled in a clinical trial evaluating vemurafenib for breast cancer. Treatment resulted in a significant reduction of metastatic lesions in the liver and peritoneum, along with a marked decrease in tumor markers, including CEA and CA15-3.

Conclusion:
This case underscores the utility of WGS in identifying rare but actionable mutations. It shows WGS's potential to expand treatment options and support personalized treatment strategies.

Discover what's possible

Uncover confident insights with configurable NGS workflow components for every step of your workflow.

Platform-agnostic workflows that were designed to support a range of application areas with seamless integration, automation, efficiency, and performance

- Whole genome sequencing
- Whole exome sequencing
- Solid tumor
- Epigenetics
- Target sequencing

Key event highlights include:

Bronze sponsor presentation:
NGS-Inspired innovations: Driving the future of gene reading

Speaker: Steven Henck, VP of NGS, Fellow
Danaher Life Sciences Division

Tuesday, February 25, 2025 • 2:40–3 pm • Calusa Ballroom

Join Steven Henck as he shares his passion for NGS and discusses the latest advancements in gene reading technologies. Learn how IDT is pushing the boundaries of what's possible and enabling groundbreaking research across the globe.

Strategic collaboration sessions

These exclusive meetings are designed to explore opportunities for collaboration, innovation, and mutual growth. Engage with our team to discuss how we can work together to drive advancements, develop custom solutions, and create impactful partnerships. Let's unlock the potential for success and innovation, together.

Join us in **Osprey 4 suite** to shape the future of genomics with Integrated DNA Technologies.

IDT™
INTEGRATED DNA TECHNOLOGIES

Informatics and Computational Biology

Extending variant detection methods for non-human species and low-coverage sequencing

Andrew Carroll[1], Daniel E. Cook[1], Lucas Brambrink[1], Juan Carlos Mier[1], Alexey Kolesnikov[1], Kishwar Shafin[1], Pi-Chuan Chang[1]

1 Google, Research, Mountain View, CA

The analysis of human genome sequencing was dramatically improved by the combination of well-characterized truth sets from NIST Genome in a Bottle, the development of deep learning methods which can learn insights directly from those truth sets, the recent demonstration of complete telomere-to-telomere genome assemblies, and the human pangenome project.

However, many of those critical resources are absent in other species. In addition, variant detection methods for humans are often optimized for higher coverage (e.g. 30x), while non-human sequencing often sequences at much lower coverage (e.g. 0.5-4x) to economically sequence populations of individuals.

Here, we demonstrate methods to train DeepVariant, a highly accurate variant calling method initially trained with human data, for these use cases. We investigate performance on a number of plant and animal species with varying genome and population structures (e.g. highly heterozygous mosquito populations, in-bred plant strains, and genomes with high repeat content or recent duplications). We characterize different training strategies, including trio-based where labels for a child are determined by a pedigree, simulated read data, and training from well-characterized parental and child lines established from later inbreeding but not a direct cross. We contrast the advantages and drawbacks of each approach, explore how much training data is needed to train effective models, and contrast the capabilities of long and short read genomics methods.

We show the ability to reduce the Mendelian Violation Rate substantially. For example, in a mosquito demonstration we reduce Mendelian Violations in a trio from 5.2 % to 0.2%. We show genome contexts where model behavior changes, for example in segmental duplication regions and high diversity regions.

For low coverage contexts, we demonstrate an increase in variant calling precision at 1x coverage from 90% to 97%, and show how this can enable detection of rare variants (e.g. 1:10,000 and lower) in low-coverage cohort sequencing which would otherwise be indistinguishable from noise.

Genome mining potential utilizing biosynthetic gene clusters of the untapped microbiome

Chandrima Bhattacharya[1,2], Jan Krumsiek[1,2,3], Christopher E. Mason[1,2,3,4]

1 Cornell University, Weill Cornell Medicine, Tri-Institutional Computational Biology and Medicine (TRI-I CBM), New York, NY
2 Cornell University, Weill Cornell Medicine, The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, New York, NY
3 Cornell University, Weill Cornell Medicine, Department of Physiology and Biophysics, New York, NY
4 Cornell University, Weill Cornell Medicine, WorldQuant Initiative for Quantitative Prediction, New York, NY

The decline in the discovery of new antibiotics, despite their crucial role in infection control, has been consistently emphasized by the World Health Organization since 2010. This issue persists even with advancements in genome mining and combinatorial chemistry, as no significant breakthroughs in novel antibiotics have emerged, further exacerbating the so-called 'silent pandemic.' Additionally, most drug discovery efforts focus on a few genera, such as Actinobacteria, limiting the exploration of vast, untapped microbial diversity. The biosynthetic machinery responsible for producing secondary metabolites (SMs), or natural products, is typically encoded by biosynthetic gene clusters (BGCs). BGCs often serve as high-throughput proxies for novel substance discovery, including antimicrobials. Recently, IMG-ABC (JGI BGC database) unveiled over 330,884 BGCs, highlighting rapid advancements in microbial genome mining. Without effective BGC prioritization algorithms, we risk continuing the post-1980s decline in antibiotic discovery due to the prohibitive cost and time required for validation.

In this study, we conducted a comprehensive, data-driven statistical analysis to develop robust methodologies for genome prioritization. We began with a curated list of 81 COGEM (Netherlands Commission on Genetic Modification) pathogenic bacteria and fungi, along with a subset of the IMG-ABC database consisting of 1,212 microbes, to identify genomic traits that could prioritize species for drug discovery. We assumed that a higher number of predicted BGCs would lead to a higher number of active BGCs, and consequently, to a greater number of small molecules that could serve as candidates for drug discovery. We found that the number of BGCs in both bacteria and fungi shows a strong positive correlation with the GC content of the species. While Actinobacteria have traditionally been the genus of choice, a previous study suggesting that species with 2.4 BGC/Mb are prolific producers aligns. We identified species from the Burkholderiaceae family and certain species from Firmicutes as promising candidates for further exploration. We discovered that bacteria and fungi produce distinct chemical classes of natural products. Our findings highlight and emphasize the importance of the biological kingdom in exploring diverse chemical classes for novel secondary metabolite discovery and provide recommendations and a strategy for computationally prioritizing microbiomes for further experimental validation.

Disease-driven changes in cellular neighborhoods as a window into spatial transcriptomics

Patrick Danaher[1], Caleb Stokes[2,3], Dan McGuire[1], Lidan Wu[1], Joseph M. Beechem[1]

1 Bruker Spatial Biology, Seattle, WA
2 University of Washington, Department of Pediatrics, Seattle, WA
3 Seattle Children's Hospital, Division of Pediatric Infectious Disease

Single-cell spatial transcriptomics poses the best kind of problem: how can an analyst possibly uncover all the biological insights contained within a dataset? Interesting biological relationships can be clearly seen, but only if the analyst knows where to look. One promising category of "places to look" are regions where a disease tissue is substantially perturbed from control tissues. To this end, we introduce "perturbation analysis", a new framework for spatial transcriptomics data analysis.

We begin by deriving a "perturbation matrix" measuring how much each gene in each cellular neighborhood is perturbed in disease vs. control samples. This alternative data matrix enables numerous analyses. First and simplest, spatial maps of gene perturbation scores are often more informative than maps of expression level. Second, we score the total perturbation of tissue regions, highlighting regions most impacted by disease, and easily flagging technical artifacts as an ancillary benefit. Third, we can identify highly perturbed genes, and we can quantify the spatial dependence of their perturbations. Fourth, by applying spatial clustering algorithms to the perturbation matrix, we can discover distinct impacts of disease across the span of tissues. Finally, clustering genes' perturbation values, we discover sets of genes perturbed in the same tissue regions as each other, uncovering the pathway biology of disease.

We demonstrate this approach in a study of mouse brains infected with West Nile virus and profiled with the CosMx® Mouse Neuroscience Panel (1000 plex). We find widespread perturbations in gene expression in brains 11 days post-infection, driven by classical antigen presentation (B2m), immune recruitment (Ccl2) and myeloid functions (C1qa/C1qb/C1qc, Cts2). At 21 days, perturbation recedes to foci with high histocompatibility 2 genes (H2-Aa, H2-Ab1) and T-cells (Cd3e). Performing spatial clustering on the perturbation matrix, we find a cluster of perturbations in the cortex at day 11; at day 21, this perturbation pattern re-localizes to the subcortical region.

In conclusion, perturbation analysis is a new scaffold for spatial transcriptomics analysis. Using it, we obtain more biologically relevant outputs from standard tools like highly variable or spatially variable gene searches, clustering genes, clustering cellular neighborhoods, or simply plotting gene expression in space.

A high-resolution human MHC reference with 282 new contiguous sequences

Julianne David[1], Liza Huijse[1], Solomon Adams[1], Joshua Burton[1], Russell Julian[1], Galit Meshulam-Simon[1], Harry Mickalide[1], Bersabeh Tafesse[1], Verónica Calonga-Solís[2], Ivan Rodrigo Wolf[2], Ashby Morrison[3], Danillo Augusto[2], Solomon Endlich[1]

1 Base5 Genomics, Inc., Mountain View, CA
2 The University of North Carolina at Charlotte, Department of Biological Sciences, Charlotte, NC
3 Stanford University, Department of Biology, Stanford, CA

Motivation:
The major histocompatibility complex (MHC) in the human genome is a gene-dense region whose genes are key to immune system function, including susceptibility to autoimmune and infectious diseases and response to cancer treatments such as transplantation and immunotherapy. Due to its extreme polymorphism and structural variation, previous contiguous sequences of the MHC region have not fully represented the breadth of its diversity. The lack of a comprehensive genomic reference hampers virtually all downstream analysis of this important region.

Results:
Here, we make a groundbreaking advance in understanding the MHC. We assembled, de novo, 282 highly accurate, fully contiguous and phased full-length MHC sequences with high population diversity using a self-learning method developed to leverage targeted or whole genome long-read data. With these sequences, we then 1) identified alleles at high resolution across 44 loci, including class I and II human leukocyte antigen (HLA) genes, discovering 1,311 high-confidence putative novel HLA allele sequences; 2) identified C4A~C4B haplotypes, finding copy number variation in the C4 genes and significant linkage disequilibrium between 5 C4 haplotypes and 14 MHC loci; and 3) built a novel "pan-MHC" reference graph that provides unprecedented accuracy for variant imputation and allele calling from short read genomic data in the previously "dark" delta polymorphic frozen block. The high population diversity of our haplotypes and graph improve the prospects for global health equity in this clinically important genomic region, and our method is extendable to other complex genomic regions.

Understanding diploid genome benchmark metrics in distinct sequencing contexts

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

1 National Institute of Standards and Technology (NIST), Material Measurement Laboratory, Gaithersburg, MD
2 Baylor College of Medicine, Human Genome Sequencing Center, Houston, TX
3 Rice University, Department of Computer Science, Houston, TX

The Genome in a Bottle (GIAB) consortium generates variant benchmarks for a set of human genomes to enable evaluation of sequencing and variant detection technologies. Despite advances, none of these methods perform equally well across the entire human genome; thus users need to make tradeoffs when selecting a variant detection pipeline. Currently, assessing these tradeoffs relies primarily on intuition regarding a method's known strengths. This challenge is further compounded by the arrival of telomere-to-telomere (T2T) diploid genomes, which enable calling more complex variants due to additional difficult regions and varying degrees of heterozygosity between haplotypes. Here we present two tools which can quantify variant calling performance in diploid human genomes.

The first is the GIAB genome stratifications, which were originally developed for GRCh37/38 but were recently extended to include the T2T-CHM13 and T2T-Q100 HG002 diploid assembly. These stratifications are bed files that denote regions of different types, such as homopolymers and difficult to map regions, which are known to present distinct challenges for each sequencing platform. These stratifications can also be used with the high-quality Q100 HG002 assembly to understand how sequencers and assemblers perform in different types of repeats. Furthermore, we added diploid stratifications to denote regions with high homozygosity, which will be useful for understanding where assemblers have errors due to erroneously collapsing non-identical haplotypes into identical haplotypes.

The second tool is called StratoMod which is an interpretable machine learning model that can produce easily understandable relationships between genomic context and predicted variant calling performance (precision or recall). We used StratoMod to predict recall using Hifi or Illumina and leveraged its interpretability to measure contributions from difficult-to-map and homopolymer regions. Furthermore, we used Statomod to assess the effect of mismapping on predicted recall using linear vs. graph-based references, and identified the hard-to-map regions where graph-based methods excelled and by how much.

We anticipate these two tools will be useful in understanding tradeoffs of emerging sequencing technologies when applied to difficult regions in diploid assemblies. While the stratifications are relatively simple and only rely on bedtools to be used, StratoMod can give much more precise predictions regarding difficulty.

Accurate isoform identification and quantification from long-read isoform sequencing with applications to bulk and single-cell Transcriptomes

Brian Haas[1], Houlin Yu[1], Dan Bartlett[1], Emily White[1], Asa Shin[1], Christophe Georgescu[1], Can Kockan[1], Akanksha Khorgade[1], James T. Webber[1], Clotilde Lagier-Tourenne[2], Michael Ward[3], Paul Blainey[1], Victoria Popic[1], Niall Lennon[1], Aziz Al'Khafaji[1]

1 Broad Institute of Harvard and MIT, Cambridge, MA
2 Massachusetts General Hospital, Department of Neurology, Boston, MA
3 National Institutes of Health (NIH), National Institute of Neurological Disorders and Stroke, Bethesda, MD

Advances in long isoform read sequencing continue to yield unprecedented insights into the functional outputs of gene expression with ever-increasing resolution. While many long isoform read sequences are full-length and sequenced from the transcriptional start site to the poly-A tail, partial read sequences stemming from incomplete reverse transcription or RNA degradation pose major challenges for unambiguous isoform identification and quantification.

We present our new computational method Long Read Alignment Assembler (LRAA) for isoform identification and quantification based on spliced genome alignments of long read isoform sequences. LRAA leverages a novel splice graph implementation, read structure annotation, compatible read collapse and artifact filtering to yield isoform structures in both reference annotation-free and reference annotation-guided modalities. In benchmarking LRAA isoform identification and quantification in comparison to 12 alternative methods using both simulated and real PacBio or ONT isoform sequences, LRAA is demonstrated to achieve top performance.

To address the inherent difficulty of benchmarking long read isoform quantification with real sequencing data regarding incomplete read sequences and lack of perfectly defined truth sets, we sequenced 2923 isoforms of 1566 human genes engineered with isoform-specific barcodes and expressed in cells. The barcoded isoforms enabled reads for incompletely sequenced isoforms to be annotated with unambiguous isoform assignments. The barcodes were then trimmed, leaving the remaining long read sequences as ground truth for evaluating quantification accuracy of tools. Leveraging these long reads, we find LRAA to be among the top-performing methods for long read isoform quantification and the most accurate method to enable both isoform identification and quantification capabilities together.

The computational efficiency of LRAA has been a great asset to us in processing and analyzing single cell transcriptomes. In our exploration of single nucleus sequencing of brain samples derived from patients with Amyotrophic lateral sclerosis (ALS), we identified disease-relevant isoforms involving cryptic splicing events evident in individual neurons. We further highlight the capabilities of LRAA for resolving isoforms and quantifying isoform and gene expression based on single cell long read isoform sequencing and related translational efforts in spatial transcriptomics.

5hmC detection in mouse tissue using long- and short-read sequencing methods

Jessica Hosea[1], Paul Hook[1], Luke Morina[1], Winston Timp[1,2]

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

The epigenome plays a key role in gene regulation in mammals and often becomes dysregulated in disease. But how epigenetic patterns are regulated themselves - especially how methylation can be lost at specific loci - is still not fully understood. Most studies of methylation in mammalian samples have focused on 5-methylcytosine (5-mC) - 5-hydroxymethylation (5-hmC) is less characterized due largely to both its low prevalence in most samples and the difficulty in its measurement. But 5-hmC is thought to be a key intermediate in demethylation as it is not recognized by DNA methylation maintenance enzymes (e.g. DNMT1). Measurement of this mark can provide important clues to how methylation is regulated.

We generated a high-coverage (~100X) dataset with both short- and long-read sequencing methods to benchmark 5-hmC across four mouse tissues with three replicates: cortex, cerebellum, liver, and heart. Despite the development of various methods for profiling 5-hmC, control samples with appreciable levels of 5-hmC are still scarce. We reasoned that biological tissues from well-characterized mouse models offer the closest approximation to a "Genome in a Bottle," with realistic levels of 5-hmC accumulation, which is rarely observed in actively dividing cultured cells.

Our analysis revealed strong correlations between the short-read enzymatic data and the long-read basecalling results, despite challenges posed by the low abundance of 5-hmC and the inherent limitations of sequencing depth. We identified genomic regions with significant 5-hmC enrichment (>10%) in windowed intervals across all tissues, using biological replicates to minimize noise. Phasing of the long-read data enabled us to detect allele-specific methylation and hydroxymethylation in these tissues.

Our study provides a valuable resource for future method development and standardization in the field of hydroxymethylation. The dataset not only validates both short- and long-read sequencing approaches but also establishes a benchmark for investigating 5-hmC's role in tissue-specific gene regulation and epigenomic modulation. We believe this work will significantly advance our understanding of 5-hmC in mammalian tissues and facilitate future epigenetic research.

TAGENE: an annotation pipeline to integrate long-read transcriptomic data into the GENCODE

Toby Hunt[1], Jose Gonzalez[1], Jonathan Mudge[1], Adam Frankish[1], on behalf of the GENCODE consortium

1 European Bioinformatics Institute (EBI), European Molecular Biology Laboratory (EMBL), Cambridge, United Kingdom

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

The availability of large quantities of accurate long-read RNA sequences from the PacBio and ONT platforms offers the potential to greatly improve GENCODE annotation. Especially, integrating these data can allow us to resolve incomplete transcript models, identify novel loci, and greatly expand our repertoire of alternatively spliced transcripts. However, the sheer scale of data availability presents a challenge to our existing annotation paradigm.

Here we discuss TAGENE, our manually supervised computational workflow for the incorporation of long-read sequences into annotation. TAGENE is designed to be highly flexible, having a modular structure, allowing us to process a variety of annotation tasks rapidly and at scale. TAGENE has first been deployed to expand our lncRNA catalog as part of the GENCODE Capture Long-read Sequencing (CLS) project, and resulted in the addition of over 130,000 novel lncRNA transcripts and ~16,000 new lncRNA genes to the Human V47 geneset, and 129,000 new non-coding transcripts and ~21,000 lncRNA genes to the Mouse M36 annotation.

We now present the next stage of this work, where we turn our attention to our protein-coding catalog. Here, a variety of projects are being used to enhance and expand our annotation, such as: expanding the 5' and 3' UTRs of existing models; resolving the exon structure of incomplete models; and identifying novel, alternatively spliced transcripts and assigning them an appropriate biotype. Preliminary results suggest that this may add up to ~181000 novel or extended transcripts in 13100 protein-coding genes.

Thus we are continuing to refine and improve the GENCODE genesets. Firstly, by incorporating new long-read transcriptomic data, and secondly by refining this expanded annotation into high-value subsets of functionally defined transcripts such as the Matched Annotation from NCBI and EMBL-EBI (MANE) set, and the newly available GENCODE Primary subset of our GENCODE Comprehensive gene annotation.

MACDADI: an open-source methylation-based classifier for pediatric CNS tumors and sarcomas for improved molecular diagnosis

Benjamin J. Kelly[1], Ke Qin[1], Lakshmi Samidurai[1], Marco Leung[1], Catherine E. Cottrell[1], Elaine R. Mardis[1]

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

Pediatric cancer is rare and presents a very distinct etiology when compared to adult cancers. Central nervous system (CNS) tumors and sarcomas, which make up approximately 30% of all pediatric tumors, can prove challenging to diagnose through clinical, radiographic, and pathologic analysis. Given the diversity of disease entities, with 100+ CNS tumor and 65+ sarcoma subtypes, molecular methods, including DNA methylation-based classification have been shown to offer diagnostic and prognostic value in these diseases. With broad training datasets including rare and common tumors, machine learning-based methylation classifiers have been developed by the community, and the World Health Organization identifies them as important molecular tools for informing diagnosis and identifying new tumor subtypes.

Use of methylation-based classifiers in clinical settings has become increasingly challenging. The most widely utilized research implementation is moving toward a commercialized model, leaving clinical sites uncertain about long-term use. Developing a classifier in-house requires a comprehensive reference dataset, and rare tumor subtypes are difficult to collect. Finally, methylation array vendors periodically update their platforms creating a potentially moving target for algorithm development and production use.

To address these concerns in such an important area, we developed MACDADI (Methylation Array Classification Diagnostic and Data Integration). The classifier is built on a regularized generalized linear model (RGLM) and trained on a subset of CpG probes present in all Illumina array platforms to ensure compatibility. The robustness of the RGLM approach enables MACDADI-CNS to classify an additional 20% of tumors versus the industry standard algorithm across 109 CNS subtypes, and MACDADI-Sarcoma an additional 24% across 65 sarcoma subtypes. In addition to classifier performance improvements, MACDADI also produces detailed clustering plots and CNV calling.

MACDADI-CNS has been clinically validated and is used to classify an average of 35 pediatric tumors per week, providing valuable and more precise diagnostic and prognostic information to clinicians, as well as informing clinical trial eligibility for children with CNS tumors across the country and world as part of the National Cancer Institute's Molecular Characterization Initiative. MACDADI-CNS is publicly available on GitHub. MACDADI-Sarcoma is currently undergoing testing with an ultimate goal of clinical validation and public release.

Training and education in the cloud with the NHGRI AnVIL

Natalie Kucher[1], Michael Schatz[1]

1 Johns Hopkins University, Baltimore, MD

As the scale of data generated and analyzed for biomedical research increases, the need also increases for accessible data and computing resources to promote open and reproducible science. Cloud-based platforms can enable resource-intensive analysis and support sharing data and tools, so it is critical to develop training to support the transition of genomic data science researchers to the cloud for both expert bioinformaticians and the next generation alike.

The NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-space or AnVIL (https://anvilproject.org/), which is a secure, cloud-based platform for storage, management, and analysis of genomic and related datasets. AnVIL integrates a number of common bioinformatics tools like Galaxy (https://galaxyproject.org/), Bioconductor (https://www.bioconductor.org/), and Workflow Description Language (WDL) workflows; so that users can bring their preferred analysis methods to the data or come in through multiple entry points for novice bioinformaticians. Instructors can use AnVIL for teaching with consistent, no-download software environments, dedicated computing resources, and participant and billing management. These advantages result in faster jobs, reduced data transfer, tracked trainee activity and spending, and flexibility for in-person and virtual training events.

We develop scalable training and outreach content which guides users in onboarding to the platform and running tutorials to learn the building blocks for cloud-computing in AnVIL. In this presentation we will showcase the available AnVIL training resources in the AnVIL Collection (https://hutchdatascience.org/AnVIL_Collection/), highlight educational modules for the Genomic Data Science Community Network (GDSCN; https://www.gdscn.org/) BioDiversity and Informatics for Genomics Scholars (BioDIGS; http://www.biodigs.gdscn.org/) research project with U.S.-based under-resourced undergraduate institutions, and the training opportunities that were presented at our inaugural AnVIL Community Conference CollaborationFests (CoFests) featuring deep learning model deployment, polygenic risk score analysis, and custom tool integration with AnVIL.

Blended Length Genome Sequencing (blend-seq): blending short and long reads to maximize variant discovery

Ricky Magner[1], Fabio Cunial[1], Sumit Basu[2], Ron Paulsen[2], Scott Saponas[2], Megan Shand[1], Eric Banks[1], Niall Lennon[1]

1 Broad Institute of MIT and Harvard, Cambridge, MA
2 Microsoft Research, Redmond, WA

We introduce blend-seq, an approach for combining short-read sequencing data with very low coverage long reads, that results in substantially improved germline variant discovery for both short and structural variants. With only 4x of long reads combined with standard short-read coverage (30x), we outperform SNP calling using short reads alone even at very high levels of coverage (60x), and improve SV recall by a factor of three across the genome while maintaining precision. For short variants, the improvements come from the few layers of long reads filling the gaps where short reads cannot easily map. We provide a detailed analysis of the additional SNPs found by our method, categorizing them by the sources of evidence that led to each discovery, compared to the NIST GIAB truth data. In the SV setting, low-pass long reads outperform short reads on their own, and by combining them with short reads we can call genotypes more accurately than with either technology in isolation. We study SV performance in several regions of interest, including difficult-to-map regions for short reads and the exome, as well as stratifying by SV type and length. We compare our SVs to the NIST Q100 HG002 assembly with curated SV calls. Our analysis confirms the expectation that blend-seq adds its greatest benefit for short variants over hard-to-map regions, and that short reads help improve SV genotypes in easier-to-map regions, where short reads map well. Our comparisons include mixtures of Illumina NovaSeq X short-read sequencing with long reads coming from either PacBio Revio or ONT R10 ultra-long reads, along with a few SV genotyping investigations done using Ultima short reads.

Structural variant detection using image-based machine learning and on-flowcell proximity information

Gavin Parnaby[1], Mitchell A. Bekritsky[1], Jeff Gau[1], Vitor Onuchic[1], Mohammed Hashemi[1], Rishi Verma[1], Pascal Grobecker[2], Gery Vessere[1], Arun Subramaniyan[1], Rami Mehio[1], James Han[1]

1 Illumina Inc., San Diego, CA
2 Illumina Inc., Cambridge, Cambridgeshire, United Kingdom

Whole genome sequencing (WGS) technologies have been widely adopted due to high accuracy and competitive pricing. However, comprehensively characterizing the entire structural profile of individual genomes has been difficult due to challenging genomic regions such as long repetitive regions and segmental duplications.

Illumina has developed a new simplified on-flowcell library preparation method on the NovaSeq™ X which delivers an expanded genome by supplementing standard sequencing reads with long range flowcell proximity information. This information provides novel insights into the genome – especially for rearrangements in repetitive and hard to map regions that have previously been challenging for standard WGS. Flowcell proximity information enables placement of reads around segmental duplications; phasing variants over long distances; and detection of complex structural events. This technology produces significant gains in small variant accuracy and haplotype phasing with median NG50 of approximately 6 Mbp when using a high molecular weight DNA extraction.

On-flowcell proximity technology physically associates reads that originated from the same input DNA molecule, providing long range information between two distant regions of the genome. Flowcell proximity information is used to visualize a genome's structure relative to a reference. Structural variants leave a unique visual signature on this image that depends on the structural variant type (i.e. insertions, deletions, inversion, etc.) and size of the event, which can be interpreted visually. In addition, automated ML-based image analysis algorithms detect and label deletions, insertions, duplications, translocations, inversions and complex SVs. Breakpoint precision down to basepair resolution is available through short read analysis.

We demonstrate our ability to accurately call large rearrangements including clinically important F8 inversions and ring chromosomes in real samples. These balanced events in low mappability regions and overlapping segmental duplications have been the most difficult to call accurately with WGS and other sequencing technologies. We visualize and highlight complex structural variation, providing new insights into genomic structure.

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

Cell-type quantification with rapid and affordable single- molecule methylation

Luke Morina[1], Courtney Hall[1], Jessica Hosea[1], Julian Rocha[1], Winston Timp[1,2]

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

While genomic studies can often rely on bulk tissue analysis to provide meaningful insights, epigenetic studies are more complex. Methylation analysis is often limited by the heterogeneous nature of tissue, with a complex distribution of cell types within a tissue sample. This can lead to either differences in cell type proportion incorrectly identified as sample-specific differentially methylated regions (DMRs). Alternatively, differences in methylation within a subset of cells can be masked by the bulk signal. Most often samples only have a gross characterization of their tissue of origin as characterizing the cell type proportions requires either microscopy paired to pathology or immunofluorescence, flow cytometry, or single cell sequencing methods, all of which are expensive and time consuming.

Methylation can be used to estimate the cell type proportions in a sample, but historically has required either sequencing or microarray methods, both of which have required high capital investment and trained personnel with complex methods. In contrast, by leveraging portable, affordable, single-molecule sequencing coupled to adaptive sampling enrichment methods, we can interrogate the heterogeneous distribution of cellular states in a tissue sample for minimal cost and time.

To achieve this, we have developed a workflow to estimate cell-type proportion with nanopore sequencing. We are validating this approach by applying it to human peripheral blood mononuclear cells (PBMCs) sequenced with nanopore sequencing along with single-cell RNA (scRNA) to validate cell-type proportions. By computationally targeted cell-type DMRs with adaptive sampling, we can focus our sequencing at only those regions informative for cell type, allowing high coverage and hence sensitivity. Ultimately, we hope to apply these strategies to single-molecule sequencing data to improve the viability of cell-type-aware multiomics. These strategies are especially critical for already extracted DNA samples which are the majority of most cohorts.

Benchmarking Lancet2 on cancer cell lines with enhanced two-tech truthsets

Rajeeva Lochan Musunuri[1], Bryan Zhu[1], Wayne E. Clarke[1,3], Timothy Chu[1], Jennifer Shelton[1], Dickson Chung[2], Shreya Sundar[2], Adam M Novak[2], Benedict Paten[2], Nicolas Robine[1], Giuseppe Narzisi[1]

1 New York Genome Center, New York, NY
2 University of California Santa Cruz, Genomics Institute, Santa Cruz, CA
3 Outlier Informatics Inc., Saskatoon, Saskatchewan, Canada

One of the challenges in cancer genomics is the ability to accurately detect somatic mutations in heterogeneous tumors, and precisely determine their clonal origin and evolution. This fundamental knowledge is central to the discovery of cancer therapies. In recent years, reductions in the cost of whole-genome sequencing have enabled researchers to address these questions in unprecedented detail. However, indels and genomic rearrangements are more challenging to detect and inspect with traditional read alignment tools since they are sufficiently different from the linear reference genome. Genome graph structures are becoming increasingly popular to encode the genomic sequences from multiple related samples, but to date these developments have been limited to the analysis of germline variants.

To address these shortcomings, we developed Lancet2 (https://github.com/nygenome/Lancet2), the successor of Lancet, a somatic variant caller which leverages local assembly and joint analysis of tumor-normal paired data using region-focused colored de Bruijn graphs. Lancet2 is a complete redesign with focus on improved performance, software maintainability, and genotyping accuracy. Due to these refactoring efforts and the use of a pull-based reactive multithreading approach, Lancet2 achieves a 10x improvement in runtime performance compared to Lancet1 and a near optimal CPU scaling making it an ideal choice for cloud deployment. We performed extensive benchmarking using multiple matched tumor/normal pairs (COLO829, HCC1143, HCC1187, HCC1395) which we sequenced at high-depth using a combination of short (Illumina) and long (ONT) reads. In comparison with other state-of-the-art callers (MuTect2, Strelka2, VarNet, DeepSomatic) Lancet2 shows the highest precision without significant loss in sensitivity.

We will present our recent results, highlighting: (1) improved genotyping approach based on the multiple sequence alignment of assembled haplotypes and re-alignment of reads to accurately calculate support; (2) optimized variant filtering with explainable glassbox machine learning models using InterpretML (https://interpret.ml); (3) use of ONT long-reads to validate variants and enhance the existing truths sets of the cell-lines used for benchmarking; (4) integration with SequenceTubeMap (https://github.com/vgteam/sequenceTubeMap) to allow the inspection of somatic variants using elegant tube-map visualizations which show read support aligned to the Lancet assembly graph.

An improved Genome in a Bottle variant benchmark based on the complete, diploid assembly of HG002

Nathan D. Olson[1], Nancy F. Hansen[2], Nathan Dwarshuis[1], Jennifer McDaniel[1], Justin Wagner[1], Adam M. Phillippy[2], Justin M. Zook[1], T2T Q100 consortium, and the Genome In A Bottle Consortium

1 National Institute of Standards and Technology (NIST), Material Measurement Laboratory, Gaithersburg, MD
2 National Institutes of Health (NIH), National Human Genome Research Institute (NHGRI), Center for Genomics and Data Science Research, Bethesda, MD

The Genome in a Bottle (GIAB) consortium aims to develop well-characterized human genome reference materials (RMs) for validating genome sequencing and variant calling methods. GIAB has produced variant benchmark sets composed of high-confidence variant calls and regions confidently identified as homozygous reference or as high-confidence variants. Traditionally, these benchmark sets are developed by integrating multiple variant callsets. However, this process is limited in its ability to characterize more complex genomic regions. Leveraging the available sequencing data and extensive characterization, the T2T consortium Q100 subgroup selected HG002 for the development of the most accurate and complete diploid genome assembly possible. We used the HG002 Q100 assembly to generate the first whole genome diploid assembly-based small and structural variant benchmark sets. These benchmark sets include regions of the genome excluded from previous benchmark sets, e.g., large indels, longer homopolymers and tandem repeats, gene conversions, and more complex SVs. The small variant benchmark set includes ~ 89.8% (2.74 Gbp) of the HG002 assembly with >3.6 million SNVs and ~950,000 indels (<50bp), >700,000 more variants than the previous v4.2.1 benchmark. The structural variant benchmark set covers nearly 90.3% (2.75 Gbp) of the HG002 assembly and nearly 30,000 variants. This represents a significant improvement over the previous v0.6 HG002 SV benchmark, which covered only 2.51Gbp and included ~10k variants. The draft benchmark sets are currently undergoing external validation to ensure they adhere to the RIDE (Reliable Identification of Errors) principle. While the GIAB variant benchmarks are a valuable resource for the community, there are limitations to variant benchmarks, particularly for complex variants in repeat-rich regions of the genome, such as recent segmental duplications that are not compatible with current variant representation standards or variant comparison methods. The HG002 Q100 assembly is enabling the development of new benchmarking approaches that are inclusive of all regions of the genome.

Analyses of extensive sequencing data of the Genome in a Bottle tumor cell line with matched normal for the development of the

Vaidehiben Patel[1], Jennifer McDaniel[1], Nathan D. Olson[1], Justin Wagner[1], Hua-Jun He[1], Zhiyong He[1], Ayse Keskus[2], Asher Bryant[2], Chunlin Xiao[3], Andrew Liss[4], Mikhail Kolomogrov[2], Justin M. Zook[1], Genome in a Bottle Consortium

1 National Institute of Standards and Technology (NIST),, Gaithersburg, MD
2 National Institutes of Health (NIH), National Cancer Institute, Cancer Data Science Laboratory, Bethesda, MD
3 National Institutes of Health (NIH), National Center for Biotechnology Information (NCBI), National Library of Medicine (NLM), Bethesda, MD
4 Massachusetts General Hospital, Department of Surgery, Boston, MA

Here, we describe a project from the Genome in a Bottle (GIAB) consortium focused on characterizing genomic measurements from a tumor cell line and matched normal tissues with explicit consent for sharing of their genomic data. This data will be used for downstream analysis such as variant calling in development of somatic variant benchmarks. Specifically, we detail the quality control and data management practices for this data as well as our work towards developing an initial somatic structural variant benchmark. Most tumor data is from a large batch of cells from the first tumor cell line (HG008-T) developed from a pancreatic ductal adenocarcinoma (PDAC) sample. This dataset, which also includes sequencing of matched normal pancreatic and duodenal tissue, currently covers: Bulk PCR-free WGS using Illumina, Element, PacBio Onso, Ultima, PacBio HiFi, and ONT; Bulk Hi-C using Phase Genomics and Arima; Single-cell WGS using BioSkryb with Illumina and Ultima; Optical Mapping using Bionano; and Cytogenetic analysis using karyotyping and directional genomic hybridization. The NIST-GIAB team performed pre-processing quality control on the sequencing data then aligned reads to GRCh37, GRCh38-GIABv3, and CHM13 followed by alignment QC before making reads and alignments publicly available on the NCBI hosted GIAB FTP site. For the alignment and QC process, we used an array of utilities including an internal snakemake workflow, the GA4GH QC pipeline for short-read data, along with long-read QC tools such as Cramino and mosdepth to produce metrics for validation of the data. The GIAB FTP site hosts raw sequencing data and alignment files along with comprehensive READMEs that comply with FAIR data principles for accessibility and reusability of the data. A preprint describing this data collection is available at https://www.biorxiv.org/content/10.1101/2024.09.18.613544v2.

To develop somatic variant benchmarks for HG008-T, the GIAB consortium has an open-working group initially focusing on structural variants and copy number aberrations. As such, this group uses IGV and Genome Ribbon to curate a set of 155 SVs including 32 translocations, 53 deletions, 52 duplications, 17 insertions, and 2 inversions that are supported by multiple technologies and/or multiple callers.

Improved variant calling via personalized latent breakpoint graphs

Victoria Popic[1], Megan Le[2], Lillian Zhang[2], Can Kockan[1], Baris Ekim[2], Houlin You[1], Brian Haas[1], Aziz Al'Khafazi[1], Bonnie Berger[2]

1 Broad Institute, Cambridge, MA
2 Massachusetts Institute of Technology (MIT), Cambridge, MA

Systematic reference bias and read alignment artifacts impair downstream calling of both small and structural variants (SVs). Although recent work has turned to pangenomes to improve read alignment and variant discovery, pangenomes both include variants absent from a given sample, as well as lack sample-specific variants. Thus, they cannot fully mitigate reference-specific bias and alignment errors. Here, we introduce personalized latent breakpoint graphs (LBGs), which capture latent structural rearrangement breakpoints based on evidence derived directly from the reads of a given sample. At a high level, similar to pangenome variation graphs, nodes in an LBG represent genome segments, while the edges represent segment adjacencies. Unlike in pangenome graphs, however, LBG edges encode latent novel sequence adjacencies discovered directly from sample reads using discordance metrics reported by standard read aligners, as well as a novel paired-subread alignment scheme, which we use to rapidly probe pairs of consecutive read fragments for discordance (non-contiguity in distance, orientation, or order on the reference genome). We show that aligning reads to an LBG constructed from a standard linear reference genome can significantly improve variant calling accuracy across multiple variant classes, variant callers, sequencing technologies, and read depths. We find that LBGs especially boost the recall of structural variants, such as inversions and duplications, reported by state-of-the-art long-read SV callers (e.g. Sniffles2), as well as decrease the rate of false positive SNPs and indels reported by state-of-the-art small variant callers (e.g., DeepVariant). Importantly, given their compatibility, LBGs can also be constructed from a pangenome reference, boosting the discovery of rare and de novo structural variants not captured in the initial pangenome graph.

Beyond methylation beta values: single-molecule long reads for methylation entropy as a biological biomarker

Ofri Ranen[1], Jonathan Jeffet[1], Sapir Margalit[1], Yuval Ebenstein[1], Yael Roichman[1]

1Tel Aviv University, School of Chemistry, Faculty of Exact Sciences, Tel Aviv, Israel

DNA methylation is recognized as a tissue-specific marker for cell of origin analysis as well as a major modulator of gene expression in health and disease and particularly in cancer.

Array and short-read sequencing commonly measure the local degree of methylation as the average methylation value of a specific methylation site. The level of methylation termed the methylation Beta value, ranges from zero to one and reflects a population average without retaining information regarding the contribution of individual DNA molecules or the correlation between neighboring methylation events. Here, we explore the potential of long-read methylation analysis to infer other methylation metrics that rely on the methylation patterns along individual DNA molecules. We examine various long-read methylation metrics provided by nanopore sequencing and optical genome mapping. These technologies provide methylation patterns along long individual molecules and allow us to test the relation between depth of coverage and the accuracy of methylation entropy assessments. We find that traditional entropy measurements derived from short reads and derived from binning sequencing data with constant bin size are highly error-prone and suggest alternative approaches that provide reliable entropy profiles for data with a common read depth of 30x-60x. We use these metrics to infer differential methylation entropy regions between different cell types and between tumors and matched adjacent tissue for mixture deconvolution.

Benchmarking 13 tools for mutational signature attribution, including a new and improved algorithm

Nanhai Jiang[1,2], Yang Wu[1,2], Steven G. Rozen[1,2,3*]

1 Duke University and National University of Singapore (Duke-NUS) Medical School, Centre for Computational Biology, Singapore
2 Duke University and National University of Singapore (Duke-NUS) Medical School, Programme in Cancer and Stem Cell Biology, Singapore
3 Duke University, Department of Biostatistics and Bioinformatics, Durham, NC

Mutational signatures are characteristic patterns of mutations caused by endogenous mutational processes or by exogenous mutational exposures. Much research has focused on benchmarking the problem of discovering mutational signatures as latent variables in somatic mutation data from multiple tumors. By contrast, there has been little benchmarking of the problem of determining which signatures are present in a given sample and how many mutations each signature is responsible for. This problem is referred to as "signature attribution". We show that there are often many combinations of signatures that can reconstruct the patterns of mutations in a sample reasonably well. Thus, human judgement incorporating multiple lines of evidence is often needed to select the most plausible attribution. We benchmarked the accuracy of 13 approaches to signature attribution, including a new approach we call Presence Attribute Signature Activity (PASA), on large synthetic data sets. These data sets recapitulated the single-base, insertion-deletion, and doublet-base mutational signature repertoires of 9 cancer types, with 2,700 spectra total across the 3 mutation types. For single-base substitution mutations, PASA and MuSiCal outperformed other approaches on all the cancer types combined. The ranking of approaches varied by cancer type, with PASA ranked 1st in 4 cancer types and MuSiCal ranked 1st in 3 cancer types. For doublet-base substitutions and small insertions and deletions, the rankings were more stable, with PASA outperforming other approaches in most, but not all of the nine cancer types. In general, the ranking of approaches varied by cancer type, and no approach achieved both high precision and recall. We believe these observations reflect the inherent challenges in signature attribution. The rankings presented will help users select approaches for mutational signature attribution and interpret results. For developers, the synthetic data offer a resource for improvement of mutational signature attribution software.

The NIH Comparative Genomics Resource (CGR): maximizing the impact of research organisms to human health

Valerie A. Schneider[1], Nuala A. O'Leary[1], Vamsi Kodali[1], Sanjida H. Rangwala[1], Françoise Thibaud-Nissen[1], Preye Akuiyibo[1], Aron Marchler-Bauer[1], Terence D. Murphy[1]

1 National Institutes of Health (NIH), National Library of Medicine (NLM), National Center for Biotechnology Information (NCBI), Bethesda, MD

NCBI is developing the NIH Comparative Genomics Resource (CGR) to maximize the impact of eukaryotic research organisms and their genomic data to biomedical research. Because all biological processes rely on genes, proteins, and the biochemical pathways they belong to, and these fundamental elements are shared across the tree of life, a comparative genomics approach can inform nearly any aspect of biology and help identify new model organisms that may be used to address issues in human health. However, the relevant data and applications for the growing number of sequenced organisms are spread across resources, challenging their findability and use.

CGR facilitates reliable comparative genomics analyses for all eukaryotic organisms through community collaboration and an NCBI genomics toolkit. The toolkit offers high-quality genomics-related resources impactful to users' research, including interconnected databases with access points enabling seamless navigation of NCBI content and interoperable data and tools that can integrate into users' workflows. We will present our recent developments on several of these toolkit resources, including MCGV, a new web-based graphical tool supporting the visual comparison of many assemblies at a single time in a single interface, expansion of functional data facilitating comparative analysis including gene ontology and expression information, and a containerized gene annotation pipeline suitable for most metazoa and plants. We will also share new features in NCBI Datasets, which provides web and programmatic interfaces to search, browse, and download genome, gene, and ortholog-related data, including new data endpoints for taxonomy, expanded orthology content, and plans for a new gene resource. We will also demonstrate how connecting community-supplied genome-related content to the organism-agnostic CGR can make it easier for researchers to use data from sources and organisms they may not have known about, and how community feedback has impacted CGR development. Through CGR, we can assist researchers in using genomic data available from thousands of organisms to answer diverse questions, including those involving human health.

This work was supported by the National Center for Biotechnology Information (NCBI) of the National Library of Medicine (NLM), National Institutes of Health (NIH).

Democratizing access to genomic data science education through the cloud

Shurjo K. Sen[1] and the Genomic Data Science Community Network

1 National Institute of Health (NIH), National Human Genome Research Institute (NHGRI), Bethesda, MD

The current genome informatics workforce does not reflect the diversity of our nation's population. Correcting this disparity requires greatly increasing the number of individuals from underrepresented backgrounds who have the necessary training to pursue careers in computational genomics and data science (CGDS). Currently, CGDS educational opportunities are primarily limited to resource-rich institutions with access to high-performance computing clusters. In 2020, NHGRI established the Genomic Data Science Community Network (GDSCN) to work with students and faculty at limited-resource institutions to develop and facilitate access to instructional materials for CGDS teaching. By supporting the development of training materials by faculty at institutions with less resources, and increasing dissemination and outreach activities, NHGRI seeks to enable a broader spectrum of undergraduate institutions to create a wider funnel at the early stages of a future CGDS workforce.

The GDSCN organizers have worked for the past two and a half years to diversify the genomic data science research community by working with 25 faculty from Historically Black Colleges and Universities (HBCUs), Hispanic Serving Institutions (HSIs), Tribal Colleges and Universities (TCUs), and Community Colleges (CCs) geographically distributed across the United States. In spite of the global pandemic, this network has built a strong community through virtual symposia, working groups, and in-person symposiums; tailored curricula using cloud-based resources that engage students in genomic analysis; and formed scientific collaborations that include manuscripts and grant proposals. Moving forward, continuity of effort is crucial to sustain momentum and build upon the hard-earned trust among the network. Towards this goal, the community has launched BioDIGS, a collaborative urban soil microbiome research project involving GDSCN network faculty and students. This project will allow us to scale up our current pilot project to reach students from diverse GDSCN institutions to participate in all aspects of research from sample collection through data analysis. Success here will result in multiple manuscripts that will involve GDSCN faculty and students as co-authors and deepen connections amongst the network. Collectively, these efforts will provide a capstone research & education experience for GDSCN faculty and students, and will have major impacts on their future educational experiences and careers.

Pangenome-aware DeepVariant

Kishwar Shafin[1], Mobin Asri[2], Pi-Chuan Chang[1], Juan Carlos Mier[1], Jouni Sirén[2], Andrew Carroll[1], Benedict Paten[2]

1 Google Inc, Mountain View, CA
2 University of California Santa Cruz, UC Santa Cruz Genomics Institute, Santa Cruz, CA

Large-scale genomics datasets derived from the human population provide valuable prior knowledge and insights for subsequent analyses. One notable example of such a dataset is the human pangenome reference that was recently released by the Human Pangenome Reference Consortium (HPRC). This pangenome consists of 94 assembled haplotypes of genetically diverse individuals. By utilizing the pangenome reference instead of a single linear reference, the accuracy of findings in various analytical stages has improved, especially in read mapping and genotyping structural variants. However, further exploration is still essential for maximizing the utility of the pangenome. DeepVariant, known as a reliable and versatile variant caller, has the potential to leverage pangenomic information for optimizing its performance. DeepVariant takes read mappings to a reference genome and then creates images of read pileups around the spots that may contain true variants. It then employs a Convolutional Neural Network(CNN) to associate the pileup images with the genotype probabilities. In this work we propose a pangenome-aware DeepVariant by augmenting the read pileup images with the HPRC haplotypes. This augmentation allows the model to utilize the pangenome as a guide for differentiating genuine signals from spurious ones within the read alignments. Our results show that using pangenome-aware DeepVariant for short-read variant calling reduces the total number of errors by up to 11%, from 29,065 to 25,781. The improvements are more pronounced when models are trained only on read alignments created by the graph aligner, vg. For vg-only models, using pangenome-aware DeepVariant reduces the total number of errors by ~30%, from 20,359 to 14,062. These improvements underscore the importance of using pangenomes for enhanced accuracy in variant calling, which impacts genomics research and clinical applications.

Comparative benchmarking of spatial transcriptomics platforms: insights from mouse brain and human FFPE tissue samples

Ashish Shelar[1], William Renock[1], Bony de Kumar[1], Shrikant Mane[1]

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

Spatial transcriptomic technologies are powerful tools for investigating targeted gene expression patterns at single-cell resolution within tissues while maintaining their spatial context. Although many new spatial technology platforms are available for research, comprehensive comparative studies of these platforms are still lacking. This study compares two imaging-based spatial transcriptomics technologies, Xenium and CosMx, and a third bead-based technology, Curio. We evaluated the performance of these platforms using fresh frozen mouse brain samples in replicates and human FFPE tissue microarrays.

We observed high correlations in gene expression profiles between replicates within the same platform. Additionally, all three platforms demonstrated high concordance in detecting common genes across panels. In conclusion, this examination of three spatial genomics technologies provides valuable insights to guide researchers in selecting the most appropriate tool for analyzing valuable samples. This study delivers essential technical and biological knowledge, aiding researchers in making informed decisions as the field evolves rapidly with advancements in imaging and larger gene panel sizes.

Using 3' single-cell sequencing to study patient-derived tumor cells reveals enhanced potential to study differential splicing at the cell-type level

Assaf Wool[1], Gila Lithwick Yanai[2*], Sergey Nemzer[1], Zohar Shipony[2], Eti Meiri[2], Masha Frenkel[1], Zoya Alteber[1], Zurit Levine[1], Eran Ophir[1], Sarah Pollock[1], Roy Granit[1]

1 Compugen Ltd., Computational Discovery, Holon, Israel
2 Ultima Genomics Inc., Fremont, CA
*Presenting author

We conducted a comparative study involving seven tumor samples from cancer patients to assess the qualitative differences between sequencing technologies for single-cell RNA sequencing (scRNA-seq). The samples were initially processed with 10x Genomics Chromium scRNA-seq, with libraries sequenced on either the Illumina NovaSeq or converted for sequencing on the Ultima Genomics UG100 platform.

Our analysis found high consistency in single-cell expression levels across the two platforms (Pearson Correlation, R = 0.98). However, there is a key difference in read alignment, since Illumina reads are paired-end, while UG100 reads are single-ended: UG100 reads consistently initiated from the 3' end of mRNA transcripts, whereas reads from Illumina libraries did not consistently reach the 3' ends. This unique feature of UG100 enables clearer demarcation of transcript termination. Leveraging this distinct characteristic, we developed a computational pipeline that harnesses the UG100 data to detect alternative splicing events and poly-adenylation alterations in thousands of protein-coding genes.

Notably, using UG100 data, we identified a truncated variant of CARD8—a core component of the inflammasome complex—suggesting a potential loss of functional activity. Further scRNA-seq analysis revealed that this truncated CARD8 isoform was reduced in macrophages, cells which are typically associated with high inflammasome activity. These findings point to a potential mechanism for regulating inflammasome function via alternative splicing.

This study emphasizes the importance of precise polyadenylation site mapping in enhancing the resolution of splicing and polyadenylation analysis at the single-cell level. Our findings offer new insights into the molecular mechanisms underlying tumor biology and immune regulation, while also providing a systematic approach to uncover additional similar events, paving the way for potential therapeutic interventions.

NCBI's EGAPx: democratizing whole-genome eukaryotic annotation

Françoise Thibaud-Nissen[1], Pooja Strope[1], Eric Tvedte[1], Vamsi Kodali[1], Eyal Mozes[1], Deacon Sweeney[1], Nathan Bouk[1], Victor Joukov[1], Pape Sylla[1], Alex Astashyn[1], Wratko Hlavina[1], Terence Murphy[1]

1 National Institutes of Health (NIH), National Library of Medicine (NLM), National Center for Biotechnology Information (NCBI), Bethesda, MD

The National Center for Biotechnology Information (NCBI) RefSeq project provides high-quality gene annotations for over 1200 eukaryotic organisms generated using NCBI's Eukaryotic Genome Annotation Pipeline (EGAP). To support the rapidly expanding needs of large-scale sequencing projects, NCBI is now developing a publicly available version of this pipeline known as EGAPx (EGAP external). EGAPx is an evidence-based pipeline using a Nextflow workflow and containerized applications to generate high-quality structural and functional annotation of most metazoan and plant genomes. The pipeline currently utilizes short-read RNA-seq and protein sequences to inform model prediction for protein-coding and lncRNA genes, with IsoSeq and ONT transcript support coming soon. EGAPx is designed for easy configuration and for execution in the cloud (e.g., AWS) or on multiple types of local HPC platforms. EGAPx annotations exhibit high sensitivity and precision when compared to gold standard annotations of human GRCh38 (GCF_000001405.40-RS_2024_08) and Drosophila melanogaster (FlyBase Release 6.54). EGAPx is being developed as part of the NIH Comparative Genomics Resource (CGR), which facilitates reliable comparative genomics analyses for all eukaryotic organisms through an NCBI Toolkit and community collaboration.
This work was supported by the NCBI of the National Library of Medicine (NLM), National Institutes of Health. The workflow and documentation for EGAPx can be found at https://github.com/ncbi/egapx.

Leveraging ultra-deep sequencing and co-assembly to resolve the bacterial, viral, and eukaryotic genomes of the coral holobiont

Braden T. Tierney[1],[2+*], Krista A. Ryon[1],[3*], James R. Henriksen[1],[4], Gabriel A. Al-Ghalith[5], David Bogumil[6], Sarah Pollock[6], Shayal Pratap[6], Nika Iremadze[6], Christopher E. Mason[1],[3],[7],[8+]

1 The Two Frontiers Project, Fort Collins, CO
2 Harvard Medical School, Department of Biomedical Informatics, Boston, MA
3 Cornell University, Weill Cornell Medicine, Department of Physiology and Biophysics, New York, NY
4 Colorado State University, Natural Resource Ecology Laboratory, Fort Collins, CO
5 Holobiome, Boston, MA
6 Ultima Genomics, Inc., Fremont, CA
7 Cornell University, Weill Cornell Medicine, WorldQuant Initiative for Quantitative Prediction, New York, NY
8 Cornell University, Weill Cornell Medicine, Institute for Computational Biomedicine, New York, NY
*co-first author
+co-corresponding author

Abstract
Untargeted metagenomics has revolutionized our view of complex microbial ecosystems, from human gut microbiota to plant root microbiomes. However, coral metagenomics has generally lagged behind other fields, often relying on older, lower-resolution methods (e.g., 16S rRNA sequencing) to characterize bacterial and other microbial communities present on reefs. This is largely due to the complexity of the coral holobiont: each polyp represents a complex amalgam of genomic information, dominated by (>90%) eukaryotic genomes (e.g., the coral, the symbiotic algae), with only a minor fraction of the DNA deriving from bacteria and viruses. Therefore, targeted, amplification-based methods for sequencing have become the preferred method for data generation and analysis, but as a result, the metagenomic architectures and gene landscape of corals are wildly underappreciated. Here, we take advantage of the new UG100 sequencing platform to demonstrate the power of untargeted, ultra-deep (hundreds of millions of reads) sequencing in resolving the dark matter of the coral holobiont. By combining deep sequencing with advanced bioinformatic co-assembly approaches, we have built a pipeline for single-shot identification of coral genomes, algal genomes, bacterial Metagenome-Assembled-Genomes (MAGs), and viral genomes. We demonstrate that with these methods, we can build databases of coral holobiont information that rival other, well-characterized metagenomic ecosystems in scale and quality. We also highlight a compendium database of publicly available coral data and coral colonies we have sequenced, providing an associated gene catalog and genomic information as a resource to the scientific community.

PATTY: a bias correction method for bulk and single-cell CUT&Tag

Chongzhi Zang[1], Shengen Shawn Hu[1]

1University of Virginia, Department of Genome Sciences, Charlottesville, VA
2University of Virginia (UVA) Comprehensive Cancer Center, Charlottesville, VA

Accurate detection of transcription factor (TF) binding sites and histone modifications (HM) on the genome-wide scale is essential for studying epigenome functions and transcriptional regulation. Cleavage Under Targets & Tagmentation (CUT&Tag) is a low-cost, easy-to-implement epigenomic profiling method that can be performed on a small number of cells and at the single-cell level. CUT&Tag experiments use the hyperactive transposase Tn5 for tagmentation. However, we find that Tn5 exhibits intrinsic sequence insertion bias. Furthermore, preference of Tn5 insertion toward accessible chromatin influences the distribution of CUT&Tag reads. Such intrinsic sequence bias and open chromatin bias can significantly confound both bulk and single-cell CUT&Tag data analysis, which requires careful assessment and new analytical methods. To address this challenge, we present PATTY (Propensity Analyzer for Tn5 Transposase Yielded biases), a computational method for systematic characterization and correction of biases in CUT&Tag data. Our results show that histone modification signals detected from CUT&Tag, after bias correction, are better associated with orthogonal biological features such as gene expression. Additionally, single-cell clustering using the corrected single-cell CUT&Tag signals better reflects cell type specificity. This new computational method can enhance the analysis of both bulk and single-cell CUT&Tag data.

Low-pass whole genome sequencing (lpWGS) for diverse populations

Peng Zhang[1], Hua Ling[1], Justin Paschall[1], Beth Marosy[1], Chrissie Ongaco[1], Jessica Gearhart[1], Rayjean J. Hung[2], Christopher Amos[3], Kimberly Doheny[1]

1 Johns Hopkins University School of Medicine, Johns Hopkins Genomics, Center for Inherited Disease Research (CIDR), Baltimore, MD
2 University of Toronto, Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto, Ontario, Canada
3 University of New Mexico Comprehensive Cancer Center, Albuquerque, NM

Array genotyping followed by imputation has been the major strategy for genome-wide association studies (GWAS) for the past 15 years. It can approximate deep whole-genome sequencing (WGS) for variants with minor allele frequency (MAF) as low as 0.11% to 0.85% depending on populations. When combined with deep whole-exome sequencing (WES), it can yield more genetic association signals with a lower cost when compared to deep WGS. The Center for Inherited Disease Research (CIDR) has generated high quality GWAS array data for over 1 million samples to investigators working to discover genes that contribute to disease and mapping a genetic path to better health.

Previously, we have shown that lpWGS is a cost-effective alternative to array, with gains in sensitivity and specificity over array followed by imputation, especially in non-European populations and for low frequency variants (e.g. MAF < 5%). In this study, we increased the sample size to 5,760 study samples with previous global screening array (GSA) data, and expanded the ancestry to more than 20 populations including Caucasian, East and West Africa, Mexican, East and Southeast Asian, Central and South America. This allows us to evaluate the performance of lpWGS in different populations for rare and common variants, comparing to those from the GSA array. We included 49 study duplicates including seven trios, two of which are in genome in a bottle (GIAB). This allows us to evaluate imputation accuracy by computing concordance and mendelian errors. Given the largest publicly accessible resources, the array data will be imputed with the TOPMed reference panel and the lpWGS with the Human Genome Diversity Project (HGDP)+1kGP reference panel. In addition to the included GIAB samples, we are deep whole-genome sequencing a subset of 222 samples, which allows us to compute specificity and sensitivity of array and lpWGS to the deep WGS gold standard. With our extensive expertise on array data and this well-controlled large study, we aim to provide best practices for the performance evaluation, quality control and data cleaning of lpWGS across diverse populations.

Error correction of Ultima Genomics reads yields Q45 sequencing quality and out-of-the-box compatibility with existing downstream analysis pipelines

Julian Hess[1], Ilya Soifer[2], Doron Lipson[*2], Gad Getz[*1,3,4]

1 Broad Institute of MIT and Harvard, Cambridge, MA
2 Ultima Genomics, Inc., Fremont, CA
3 Massachusetts General Hospital, Boston, MA
4 Harvard Medical School, Boston, MA
*These authors co-supervised the work.

The scale and low cost of the Ultima UG100 make it a compelling platform for genomics applications requiring deep sequencing of large cohorts, such as cancer whole genomes. However, the unique non-terminated flow-based chemistry that facilitates Ultima's efficiency introduces error modes that are difficult for existing downstream analysis tools and pipelines to handle, which were designed around traditional terminated sequencing-by-synthesis (SBS) chemistries. The lack of terminated cycles causes indel errors to recur at homopolymers, since the sequencer does not discretely call each base within the homopolymer, which instead is called as a single unit whose length can be mischaracterized. This is in contrast to SBS, whose predominant error mode is not indels but substitutions caused by dephasing of its terminated cycles.

While special variant callers have been developed specifically for Ultima sequencing alignments and have demonstrated accuracy on par with other platforms, it is essential that sequencer output also be out-of-the-box compatible with all other downstream methods. We therefore developed a self-supervised deep-learning model to error correct Ultima data at the level of the individual read sequence, which virtually eliminates homopolymer artifacts that could be confounded with true sequence variation. Our error correction not only mitigates homopolymer artifacts but also other artifact modes (such as PCR errors from library amplification or formalin-induced deamination due to FFPE preservation), reducing the overall error rate to less than 1 artifact per 30000 sequenced bases (~Q45). Since the model is self-supervised, it does not need any explicitly labeled training data, making it easy to retrain on samples with distinct artifact modes.

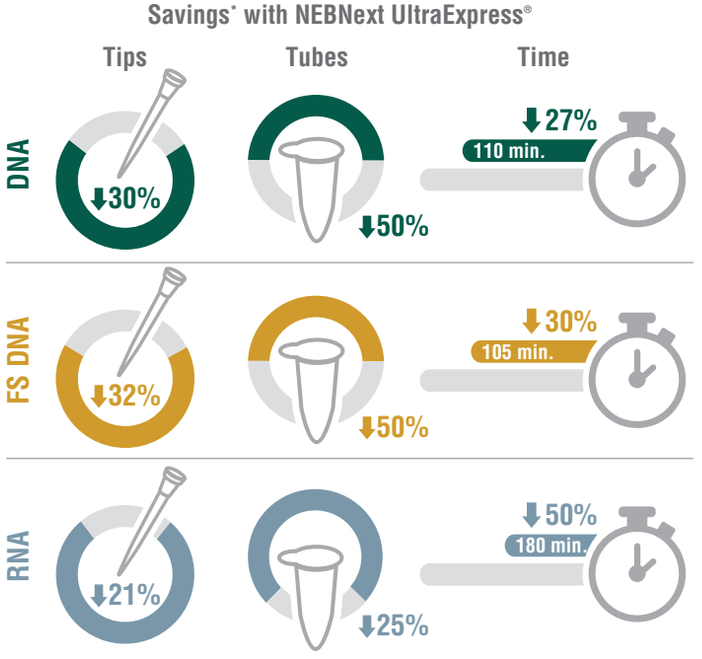
Error-corrected output is turnkey compatible with any existing downstream analyses, including well-established accurate somatic mutation detection pipelines. We ran error-corrected Ultima alignments through these pipelines and rigorously benchmarked the calls using a ground truth set of somatic mutations from well-studied cell line tumor/normal pairs sequenced on a variety of platforms, and found that performance both in terms of variant calling sensitivity/specificity and overall sequencing accuracy was not only comparable to but often superior than other platforms.

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

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

Let NEBNext UltraExpress speed up your library prep with:

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



* As compared to NEBNext® Ultra™ II

Visit New England Biolabs in **Osprey 5** to learn more.

NEW ENGLAND Biolabs®

Biology & Medical Omics (Non-Cancer)

Changes in the transcriptomic landscape of the hypothalamus in Prader-Willi Syndrome

Colleen R. Bocke[1], Megan Pater[1], Grant R. Kolar[1], Willis K. Samson[1], Gina L.C. Yosten[1]

1 Saint Louis University School of Medicine, Department of Pharmacology and Physiology, Saint Louis, MO

Prader-Willi syndrome (PWS) is a complex genetic disorder with distinct developmental stages. In infancy PWS manifests as failure to thrive, hypotonia, and difficulty feeding. During early childhood, a shift occurs that leads to hyperphagia, behavioral challenges, and endocrine disruption. Many of these challenges are rooted in hypothalamic dysfunction. Previous work on PWS has focused on clinical characterization and treatment to improve phenotypic outcome, which has benefited some patients, although the specific mechanisms that underlie many PWS symptoms remain unresolved. To date, animal models of PWS have failed to recapitulate the full scope of the disorder, adding additional challenge to mechanistic characterization of the syndrome. This project aims to address this gap in knowledge by examining the spatial transcriptomic landscape of hypothalamic neurons from individuals with PWS. While some work has been done to examine transcriptional changes in the hypothalamus of these patients through single-cell sequencing, much is left to be explored.

For this project, we employed Nanostring's CosMX Spatial Molecular Imaging (SMI) technology. CosMX SMI offers high resolution mapping of gene expression within tissues, and allows for the use of formalin fixed, paraffin embedded tissue, which greatly expands the pool of available tissue. We obtained hypothalamic tissue from individuals with PWS and age, sex, and race matched unaffected controls from tissue banks. Differential expression analysis of neurons in PWS and controls showed differences in metabolism and addiction pathways. Importantly, we found that a subcluster of CRH neurons was absent in PWS patients. The closest neuronal neighbor of these neurons was a cluster of neurons with downregulated ACTN4 expression in PWS donors but not in unaffected controls. This could indicate alterations in secretion or the cytoskeletal structure in PWS patients. These differences could potentially be influenced by changes in cellular neighborhood. These findings may reveal cellular processes that contribute to the PWS phenotype.

Elucidating the genomic underpinnings of idiopathic scoliosis with whole-genome sequencing data

Awtum M. Brashear[1], Anxhela G. Gustafson[1], Andrew Quitadamo[1], Rinu Nair[1], Kamran Shazand[1]

1 The Shriners Children's Genomics Institute, Tampa FL

Idiopathic scoliosis (IS) is defined by the presence of an abnormal structural spinal curvature of greater than or equal to 10º in the coronal plane in the absence of an underlying vertebral malformation. Approximately 3% of the school-age population is affected with IS. Most affected individuals have mild clinical features with 10% of cases requiring bracing for curves between 25-40º and surgery for curves greater than 50º. Approximately, 25% of patients with IS will fail bracing. Currently, there are no known biomarkers to help determine which patients have spinal curves that will progress to bracing and surgery, and which patients will fail bracing treatment. IS is thought to have genetically heterogeneous causes. Our goal is to identify biomarkers that can be used to improve diagnosis and guide personalized treatment approaches for our patients and the community at large.

A whole genome sequencing (WGS) effort was completed at Shriners Hospital Centers, on a cohort of 500 patients with IS. In addition to probands, WGS was obtained for parents and siblings when available. Accompanying phenotypic data was obtained by consulting clinicians. Using this dataset, we have examined the genomics of IS using cohort- and family-level analysis of Single Nucleotide Variants (SNVs), Copy Number Variants (CNVs), and Structural Variants (SVs). We have identified a key biological pathway that is suspected to be significantly perturbed in IS. We are currently pursuing a dual functional validation strategy with both multiplexed assay for variant effects (MAVE) approach to investigate the transcriptomics and a zebrafish scoliosis model where several individual and combined variants would be introduced.

Microsampling for multi-omic profiling and characterization of individuals undergoing extremely demanding physical activities and its application to marathon runners

Robert Chen[1], Erik LeRoy[1,2], Anurag Sakharkar[4], Jacqueline Proszynski[1], Ashley Kleinman[1,3], JangKeun Kim[1], Jiwoon Park[1], Mohith Arikatla[1], Natalie Peralta[1], Sara Omar[1], Christopher Mason[1]

1 Cornell University, Weill Cornell Medicine, Department of Biophysics, New York, NY
2 New York University (NYU) Grossman School of Medicine, Association of Spaceflight Professionals New York, NY
3 Massachusetts Institute of Technology, Department of Earth, Atmospheric, and Planetary Sciences, Cambridge, MA
4 University of Saskatchewan, College of Arts and Science, Saskatoon, Saskatchewan, Canada

Physical fitness is strongly associated with lower morbidity and mortality across multiple age-related comorbidities, as shown in various evidence-based medical studies. However, extensive exercise involves complex systemic physiological adaptations that are not well understood with standard retrospectively collected clinical data. Multi-omic profiling has the potential to generate multimodal insights from thousands of biomarkers measured longitudinally. We introduce a conceptual framework for decentralized multimodal data collection and subsequent multi-omic measurements. We demonstrate a proof of concept in 6 participants (3 marathon runners performing a 13.1-mile run, 3 controls).

The participants underwent data collections via blood microsampling (one Tasso+ EDTA 1mL tubes, one full Whatman 905 dried blood spot card (5 circles), wearable devices (continuous glucose monitor, smartwatch), and subjective metadata surveys at five timepoints: (R-7, R-1, mile 11 of 13.1-mile run, R+1, R+7; where R-n or R+n indicate n days before or after the run). Participants were instructed to collect all specimens independently without professional guidance for all timepoints other than the mid-run sample for athletes). User feedback uncovered a 97% success rate with Tasso+ devices, and a 81% success rate with Whatman DBS cards. On average, participants took approximately 6 minutes to collect all specimens at each timepoint, filled up 90% of the volume across all EDTA tubes (900 uL), and filled up 85% of the surface area of dried blood spot cards (200 uL). Participants generally considered the microsampling devices easy to use (1.6 of 10 points for Tasso+ tubes, 5.4 of 10 points for Whatman dried blood spot cards (10=hardest, 1=easiest)), and considered the continuous glucose monitors useful for providing individual feedback on their exercise bout (8 of 10 points (10=most useful)).

Future work includes untargeted screening via proteomics, inflammomics and metabolomics panels for hypothesis generation around adaptations to exercise, as well as comparison across larger cohorts in varied contexts under which individuals carry out physically demanding tasks (e.g., astronauts, scuba divers, race car drivers, and competitive weightlifters).

Complementary omics approaches across diverse clinical cohorts identify a biomarker of life-threatening illness from influenza, RSV, and SARS-CoV-2 infection and multi-system inflammatory syndrome

Jeremy Chase Crawford[1], Xiaoxiao Jia[2], Heather Smallwood[3], Robert Mettelman[1], Amanda Green[4], Lee-Ann Van de Velde[1], Lindsay Brown[5], Ryan Thwaites[6], Tanya Novak[7], Adrienne Randolph[7,8], Paul Thomas[1,8], Tim Flerlage[9], Shawn Campagna[5], Zhongfang Wang[2,10], Brendon Chua[2], Katherine Kedzierska[2,8]

1 St. Jude Children's Research Hospital, Department of Host-Microbe Interactions, Memphis, TN
2 The University of Melbourne, Department of Microbiology and Immunology, The Peter Doherty Institute for Infection and Immunity, Melbourne, Victoria, Australia
3 The University of Tennessee Health Science Center, Department of Pediatrics, Memphis TN
4 St. Jude Children's Research Hospital, Department of Infectious Diseases, Memphis, TN
5 University of Tennessee, Biological and Small Molecule Mass Spectrometry Core, Knoxville, TN
6 Imperial College London, National Heart and Lung Institute, London, United Kingdom
7 Harvard Medical School, Boston Children's Hospital and Department of Anesthesia, Department of Anesthesiology, Critical Care, and Pain Medicine, Boston, MA
8 Center for Influenza Disease and Emergence Response (CIDER), Atlanta, GA
9 University of Rochester Medical Center, Department of Pediatrics, Infectious Diseases, Rochester, NY
10 Guangzhou Medical University, First Affiliated Hospital of Guangzhou Medical University, Guangzhou Institute of Respiratory Health, State Key Laboratory of Respiratory Disease & National Clinical Research Center for Respiratory Disease, Guangzhou, China

Acute respiratory infections are characterized by varied disease severities and corresponding morbidity rates, and early identification of patients at risk for severe disease is integral to improving patient outcomes. We therefore set out to identify biomarkers of severe respiratory disease by studying cohorts with patients experiencing life-threatening respiratory symptoms due to acute viral infection. Bulk and single-cell gene expression assays and lipidomics were used to compare patients with life-threatening illness due to acute respiratory viral infections (including avian and seasonal influenza, SARS-CoV-2, and RSV) to healthy individuals or patients with less severe disease. Our analyses identified OLAH (oleoyl-ACP hydrolase), a typically rarely expressed enzyme involved in fatty acid biosynthesis, as a biomarker of life-threatening disease across multiple viral infections. Among adults hospitalized with A(H7N9) influenza, those who succumbed to disease exhibited early elevations of peripheral OLAH transcription that persisted throughout the course of hospitalization. Likewise, single-cell analyses showed significant increases in OLAH among adults hospitalized with seasonal influenza compared to healthy subjects, particularly in peripheral monocytes and macrophages. OLAH was also identified as a biomarker for life-threatening disease outside of influenza infection; for instance, among children hospitalized with SARS-CoV-2, OLAH was the single-most differentially expressed gene increased in patients with life-threatening respiratory dysfunction compared to those with no-to-minimal respiratory dysfunction.

(continue to page 93)

(continued from page 92)

Furthermore, lipidomics analysis of adults infected with SARS-CoV-2 showed significant up-regulation of the main catalytic products of OLAH among hospitalized patients. Single-cell analyses of samples obtained from the lower respiratory tracts of children with acute lung failure due to seasonal influenza and/or RSV infection similarly showed upregulation of OLAH in monocytes, macrophages, and neutrophils, potentially implicating this peripheral biomarker in mechanisms underlying disease at the site of infection. Analyses of samples obtained from human challenge models of multiple respiratory infections consistently demonstrated that OLAH was not elevated during mild infections. Network analysis of the underlying datasets pointed to OLAH as a potential mediator of inflammation among specific myeloid subsets that traverse the respiratory tract and the periphery. In conclusion, upregulation of OLAH transcription during viral respiratory infection correlates with myeloid-mediated inflammation and provides an early biomarker of life-threatening illness.

Comprehensive integrated single-cell and spatial multi-omics defines novel myeloid biology in ulcerative colitis

Jenny Ding[1], Yawei Li[1], Yiming Li[1], Alan Robinson[1], Cenfu Wei[1], Stephen B. Hanauer[1], Alyssa Rosenbloom[2], Kimberly Young[2], Mithra Korukonda[2], Liang Zhang[2], Kathy Ton[2], Michael Patrick[2], Prajan Divakar[2], Joseph Beechem[2], Yuan Luo[1], Parambir S. Dulai[1]

1 Northwestern University, Chicago, IL
2 Bruker Spatial Biology, Seattle, WA

A rapid growth in single-cell profiling and spatial omics' has enabled novel discovery and identification of biological insights. Gaps remain in accurately annotating cell types, data integration, leveraging the strengths of orthogonal data types to guide scientific discovery. We curated and integrated public and internal single-cell RNA sequencing datasets (over 650,000 cells) for ulcerative colitis (UC) to build a foundation atlas. Data and biology driven clustering identified 59 fine-grain clusters and a panel of highly variable and/or biological marker genes for cell type annotation. A second dataset of CITESeq was created (over 100,000 cells) to leverage orthogonal protein expression alongside 5' single-cell RNA sequencing and TCR/BCR variability to better guide RNA based inference annotations. A unique myeloid population was identified that contributed to UC pathology, and cell-cell interactions of this single-cell RNA sequencing dataset suggested unique interactions as potential determinants of inflammation. Subsequently, we completed single-cell spatial whole transcriptome (WTX) imaging using the CosMx platform (over 18,000 genes) and developed a unique dataset which demonstrated gene diversity comparable to traditional single-cell RNA sequencing. Using our reference single-cell RNA and protein-RNA datasets, we systematically benchmarked available spatial annotation packages and evaluated potential imputation approaches applied to CosMx 6K and 1K data generated from sequentially cut sections in blocks used for spatial WTX. Through these multi-modal analyses, we confirmed novel myeloid biology and further refined spatial interactions governing mechanisms through which this myeloid population contributed to inflammation. Finally, we leveraged a novel multi-omic spatial approach by combining the CosMx 6K RNA panel and the CosMx 64-protein panel, run sequentially on the same slide to generate the first ever large-scale, high-plex, and same-cell multi-omic spatial dataset for colon tissue. This further refined spatial annotations and guided the identification of unique spatial interactions in UC. To date, our work represents the most exhaustive and comprehensive single-cell and spatial dataset available for colon tissue which was used to guide accurate spatial cell type annotations and biological discovery.

Generating higher-quality transcriptomic data from challenging sample types

Nicole Francis[1], Can Kockan[1], Dan Dubinsky[1], Jason Lam[1], Cole Walsh[1], Tom Howd[1], Stacey Gabriel[1], Niall Lennon[1]

1 The Broad Institute, Broad Clinical Labs, Cambridge, MA

High-quality transcriptomic data from tissues provide valuable insights into gene expression patterns specific to various tissue types, enhancing our understanding of biological processes, disease mechanisms, and treatment effects. However, current RNA library preparation methods often exclude samples of lower quality or are prohibitively costly and inefficient. Tissues that are challenging to collect or typically yield poor-quality RNA present significant barriers, increasing costs and complicating library generation. Factors such as preservation methods, degradation over time, and the intrinsic characteristics of the tissue can contribute to diminished RNA quality. Collaborators may find it difficult to acquire optimal tissue samples and often have to work with what is available to them.

To support researchers working to advance our understanding of biology, Broad Clinical Labs evaluated products from Watchmaker's RNA portfolio to establish an automated, scalable, robust, and cost-effective bulk RNA library preparation method that accommodates tissues and RNA of varying quality. This evaluation included input titrations, quality titrations as determined by RIN/DV200 metrics, tissues sourced from diverse sites and samples with poor 260/230 ratios.

We demonstrate the ability to generate high quality sequencing libraries for tissue types that often fail through standard poly(A) methods using a Watchmaker RNA workflow. Additionally, we ran quality titrations as determined by RIN/DV200 metrics, and successfully generated high-quality cDNA libraries from samples with low to mid RIN scores (3.0 - 6.0). Overall, libraries generated from samples with lower RIN scores yielded 5x - 10x more library yield using a Watchmaker RNA workflow. Watchmaker libraries also had 60% lower duplication rates than libraries generated from samples with identical RIN scores using a standard poly(A) selection method. Here, we present the data from these evaluations, alongside proposed workflows for RNA samples of all qualities, aimed at empowering researchers to address critical biological questions.

Vaccine discovery challenges due to phase variation dynamics in non-typeable *Haemophilus influenzae*

John M. Gaspar[1]

1 Merck & Co., Inc., Data Science & Scientific Informatics, Cambridge, MA

Acute otitis media is a painful ear infection that frequently affects children and is commonly caused by one of several species of bacteria, including Non-Typeable Haemophilus influenzae (NTHi). The creation of a vaccine targeting NTHi is complicated by the presence of phase-variable genes whose expression can be rapidly and reversibly switched on or off due to the gain or loss of a short repeat in the DNA sequence. Phase-variable genes often encode immunogenic proteins, such as enzymes affecting bacterial cell surface structures, and these would be ideal vaccine candidates except for their unstable expression. Limiting vaccine discovery efforts to non-phase-variable genes may also be insufficient, since NTHi also possesses a phase-variable DNA methyltransferase that controls the expression of multiple other sets of genes.

We created a novel, simple method to quantify the dynamics of phase variation from whole genome sequencing datasets. We then applied this method to multiple strains of NTHi. In any given strain, most phase-variable loci had one particular reading frame that was dominant, but subpopulations of the two minor frames were also observed. This indicated that phase variability did not occur solely due to selective pressure but rather existed at low levels even in a steady-state population. The minor populations were more prevalent in loci with longer repeat regions, and also in genes known to be actively phase variable during infections. These results highlight the importance of avoiding genes under the influence of phase variation altogether for vaccine discovery.

Systematic analysis of the impact of short tandem repeats on gene expression

Alon Goren[1]

1 University of California, San Diego (UCSD), The Center for Neural Circuits and Behavior, San Diego, CA

We and others have identified thousands of associations between polymorphic short tandem repeats (STRs; repeat units 1-6bp) and gene expression, splicing, blood/serum biomarkers, and other complex traits. These studies have suggested multiple mechanisms by which STRs impact gene regulation, but molecular effects of STRs remain difficult to study systematically due to challenges in synthesizing, amplifying, and sequencing these low complexity regions.

Here we optimized a massively parallel reporter assay (MPRA) to characterize STRs. We first designed a library of 3-4 variants in each of 30,516 unique human promoter-proximal STRs plus 55-76bp (average 68bp) of surrounding genomic context for a total of 100,000 elements attached to random barcodes and cloned upstream of a reporter. We used two platforms, Illumina and Element avidity sequencing, to determine which elements are associated with each barcode. Compared to Illumina, Element showed a 20% increase in reads passing quality filters and resulted in detection of 4215 additional unique alleles, the majority of which consisted of homopolymers >25bp, which were not captured by Illumina.

We transfected this MPRA to HEK293T cells in triplicate which showed highly reproducible reporter expression across replicates (R=0.98). We identified 1,366 out of 19,818 loci with significant associations between repeat copy number and expression, with a strong bias toward positive effect sizes ($p<1.04e-110$, binomial distribution testing). We designed a second array to perform more detailed study of top regulatory STRs by modifying the repeat unit, orientation, and length, for a total of 200-300 perturbations per locus. This second array identified repeat unit sequence as the key driver of differences in expression associations across loci, whereas strand and sequence context had far weaker effects. Key findings include confirming strong positive effects of GC-rich repeats on expression, consistent non-linear effects of AAAC/GTTT repeats, and linear effects of AGCG and similar repeats that only have an effect after a certain copy number threshold is reached.

Overall, in this work we identified optimal conditions for studying STRs using MPRAs, allowing us to perform the first in depth interrogation of the impact of STR characteristics on transcriptional regulation and study the critical roles STRs play in genome function.

Correlative analysis of wastewater trends with clinical cases and hospitalization through five dominant variant waves of COVID-19

George S. Grills[1], Qingyu Zhan[2], Mark Sharkey[3], Ayaaz Amirali[2], Cynthia Beaver[1], Melinda Boone[1], Samuel Comerford[1], Aaron Ruby[1], H. Renata Gallegos[1], Naresh Kumar[4], Bhavarth Shukla[3], Natasha Schaefer Solle[1,3], Benjamin Currall[1], Sion Williams[1], Kristina M. Babler[2], Christopher E. Mason[5], Helena Solo-Gabriele[2]

1 University of Miami, Miller School of Medicine, Sylvester Comprehensive Cancer Center, Miami, FL
2 University of Miami, College of Engineering, Department of Chemical, Environmental, and Materials Engineering, Coral Gables, FL
3 University of Miami, Miller School of Medicine, Department of Medicine, Miami, FL
4 University of Miami, Miller School of Medicine, Department of Public Health Sciences, Miami, FL
5 Cornell University, Weill Cornell Medicine, Department of Physiology and Biophysics, New York, NY

Wastewater-based epidemiology (WBE) has been utilized to track community infections of Coronavirus Disease 2019 (COVID-19) by detecting the RNA of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) in wastewater. Correlations between community infections and wastewater SARS-CoV-2 RNA levels can change as new SARS-CoV-2 lineages emerge. This study analyzed SARS-CoV-2 RNA and indicators of human waste in wastewater from two different sized sewer sheds, the University of Miami campus and the Miami-Dade County Central District wastewater treatment plant, over five different COVID-19 variant dominant periods (i.e., Initial, Pre-Delta, Delta, Omicron and Post-Omicron). Wastewater SARS-CoV-2 RNA levels were compared to COVID-19 clinical cases and hospitalizations to evaluate correlations. Although correlations between documented clinical cases and hospitalizations were high, the correlations of clinical cases and hospitalizations for a given wastewater SARS-CoV-2 level varied depending upon the variant analyzed. The correlative relationship was significantly steeper for the Omicron-dominated period (i.e., more clinical cases correlated with higher levels of SARS-CoV-2 RNA found in wastewater). For hospitalizations, the relationship with SARS-CoV-2 RNA level was steepest for the Initial wave, followed by the Delta wave, with flatter slopes during all other waves. Overall, results were interpreted in the context of SARS-CoV-2 virulence and vaccination rates among the community. The results from this effort can inform public health strategies on local and community levels and may serve more broadly as a model for other existing and emerging pathogens.

Health genomics in ex situ populations of gorillas and other mammals

Rachel Johnston[1,2], Jenny Tung[3,4,5], Marietta Danforth[6], Elinor Karlsson[2,7], Eric Baitchman[1,2]

1 Zoo New England, Division of Animal Health and Conservation, Boston, MA
2 Broad Institute of MIT and Harvard, Vertebrate Genomics Biology Group, Cambridge, MA
3 Max Planck Institute for Evolutionary Anthropology, Department of Primate Behavior and Evolution, Leipzig, Saxony, Germany
4 Duke University, Department of Evolutionary Anthropology, Durham, NC
5 Duke University, Department of Biology, Durham, NC
6 Detroit Zoological Society, Great Ape Heart Project, Detroit, MI
7 University of Massachusetts (UMass) Chan Medical School, Program in Bioinformatics and Integrative Biology, Worcester, MA

The genetic causes of over 6,000 Mendelian diseases have been discovered in humans, due to monumental technological strides in human medicine and genomics. These advances also open new opportunities to study the genetic causes of health and disease in other species. Such work can both help improve the health and management of the study species and also shed light on the predictors, evolutionary history, and comparative genomics of disease in humans. Through the new Center for Zoonomics—a partnership between Zoo New England and the Broad Institute of MIT and Harvard—we are working to bridge the gap between health genomics and animal population genetics to improve animal management and health. As a proof of concept study, we are conducting population-level genome resequencing in captive Western gorillas to study heart disease, the leading cause of death for these animals. Results highlight the potential role of genetics for impacting gorilla health, and suggest that management of this population can be improved based on genomics-informed decision-making. This project is laying the groundwork to perform health genomics on captive animal populations more broadly.

Use case for a novel benchtop sequencer: detecting low-frequency mutations in clonal hematopoiesis

Akinori Kanai[1], Yukie Kashima[2], Kenichi Makishima[3], Mamiko Sakata-Yanagimoto[3], Yutaka Suzuki[1]

1 The University of Tokyo, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, Tokyo, Japan
2 The University of Tokyo, Life Science Data Research Center, Graduate School of Frontier Sciences, Tokyo, Japan
3 University of Tsukuba Hospital, Department of Hematology, Tsukuba, Japan

New short-read sequencers, characterized by improved read quality, are reaching the market. We have installed and started operating an AVITI sequencer from Element Biosciences, producing read quality data exceeding Q40 (error rate: 0.01 percent). We compared AVITI sequencing results with their conventional sequencer to confirm its performance in whole-genome sequencing (WGS), mRNA-seq, single-cell RNA-seq and panel sequencing. The WGS data showed less errors and higher read quality than conventional data. The comparison of SNVs showed 90 percent agreement with conventional data, and IGV visual inspection confirmed the presence of SNVs that did not match. The mRNA-Seq data showed a correlation of R = 0.99 with conventional data, confirming that the data were comparable. Human peripheral blood mononuclear cells (PBMC) were used for scRNA-Seq analysis. Both the number of analyzed cells and the median number of detected genes did not differ from the results of conventional sequencers, and the percentage of cell types was maintained. Correlations of gene expression levels per cell type were compared, and correlations exceeding R = 0.9 were obtained for cell types with sufficient cell numbers.

Thus, AVITI provides high quality data that is comparable to conventional sequencers. We took advantage of these characteristics to perform clonal hematopoietic (CH) analysis using healthy elderly PBMCs. By using Twist exome 2.0 panel and CH custom panel, we were able to confirm mutations in candidate genes that may be involved in clonal hematopoiesis. These mutation rates were low, around 1 percent, and required a high level of confidence in the reads.

AVITI has further increased read quality from Q40 to Q50 (error rate: 0.001 percent) using a new sequencing kit from Element called Cloudbreak UltraQ. We would like to share the status of these new sequencers.

Imprinting disorders and methylation dysregulation: the significance of methylation outliers

Rhina Kaur[1], Tanner Jensen[2], Devon Bonner[2], Jon Bernstein[2], Matthew Wheeler[2], Stephen Montgomery[2]

1 Howard University College of Medicine, Washington, DC
2 Stanford University School of Medicine, Stanford, CA

In autosomal inheritance, offspring receive two complete sets of chromosomes, one from each parent. However, in imprinting genes—approximately 1% of the human genome—only one allele is expressed while the other is epigenetically silenced. Silencing of imprinting regions, primarily through DNA methylation, leaves only one functional copy of the gene from a single parent. Disruptions in the imprinting process can lead to distinct congenital diseases called imprinting disorders. Our investigation explored whether patients with methylation abnormalities in known imprinting genes have phenotypes that overlap with the clinical manifestations of the linked imprinting disease. After curating a list of imprinting genes and associated disorders, we employed a pipeline to identify methylation outliers within these regions using long-read sequencing data from the GREGoR Consortium U06 release including 99 PacBio and 95 Oxford Nanopore genomes. Our analysis revealed 12 samples, of which 8 were affected individuals, with significant deviations in methylation levels in the imprinted loci of interest. Notably, a patient we identified as a loss of methylation outlier in the imprinted GNAS gene was clinically diagnosed with Pseudohypoparathyroidism Type 1B, validating our approach. Other cases also demonstrated strong phenotypic overlap with expected imprinting disorders, although we encountered significant bias in haplotype coverage when mapping methylation in many of these samples. This may be due to lower sequencing depths, confounding our results. Despite this limitation, the observed clinical overlap emphasizes the potential of this analysis for diagnosing imprinting diseases. Future improvements such as a haplotype-aware analysis that can correct for and annotate biased haplotype coverage could significantly enhance our ability to uncover imprinting disorders in patients who face prolonged diagnostic uncertainty.

Multi-omic profiling in human recipients of pig xenografts

Brendan J. Keating[1,2,3], Eloi Schmauch[4], Brian Piening[5], Simon Williams[1,2,3], Qian Gao[1,2,3], Mike Snyder[6], Adam Griesemer[1,2], Jef Boeke[1,2,3], Robert A. Montgomery[1,2]

1 New York University (NYU) Langone Health, NYU Langone Transplant Institute, New York, NY
2 New York University (NYU) Grossman School of Medicine, Department of Surgery, New York, NY
3 New York University (NYU) Langone Health, Institute for Systems Genetics, New York, NY
4 Broad Institute of MIT and Harvard, Cambridge, MA
5 Providence Cancer Center, Earle A. Chiles Research Institute, Portland, OR
6 Stanford University, Department of Genetics, Palo Alto, CA

Introduction: Only 1 in 4 people on US transplant waitlists will received an organ. Since 2021 gene-edited pig xenografts have been transplanted into living and deceased humans with varying success.

Methods: We performed 6 pig xenograft to humans xenotransplants: (a) pig kidney to human decedents (2 procedures for 3 days, and one for 61 days); (b) 2 pig heart to decedents (both for 3 days) and (c) pig kidney to a living human (47 days). Blood samples were obtained from decedent every 6 hours (for 3 day procedures), or everyday (61 day procedure), with living human samples available every 3 days. Over 30 tissue biopsies were available across the 6 procedures including a biopsy confirmed antibody mediated rejection (AbMR) event midway through the 61-day procedure. We performed WGS, plasma lipidomics, metabolomics, proteomics (Seer and Alamar) and cell-free DNA screening on >140 timepoints along with scRNAseq, B- and T-cell repertoire seq, bulk RNAseq in >60 PBMC timepoints, with >15,500 virus strains/bacteria screened and in situ and/or geospatial transcriptomics on 18 biopsy timepoints (Visium and Xenium, 10x genomics), across rejection, infection and treatment events.

Results: In the 61-day procedure >6,870, and 1,150 proteins were assignable to human and pig proteomes with >90% of all complement proteins detectable. Longitudinal signatures correlate with the AbMR and the subsequent successful treatment. scRNAseq showed concerted immune-cell responses preceding AbMR and Xenium data showed >20% of all cells in the AbMR timepoint are human infiltrating immune cells, with a sharp drop-off of human cells after AbMR treatment. In the first pig heart to human decedent procedure massive T-cell expansion on a background of ischemia reperfusion injury was evident across all omics after 36 hours which were not evident using conventional histology. In our living human a significant number of infections were evident from genomic screening including a number which were not detected using conventional infectious disease screening, and correlated with BCR/TCR findings.

Conclusion: Multiomic profiling of blood and tissue samples in pig to human xenotransplants reveals yields a number of immunological insights including a successfully treated AbMR episode and infection.

Chromatin changes associated with neutrophil extracellular trap (NET) formation reflect environment

Justin Cayford[1], Brandi Atteberry[1], Akanksha Singh-Taylor[1], Benjamin P. Berman[2], Theresa K. Kelly[1]

1 Volition America, Innovation Lab, Carlsbad, CA
2 The Hebrew University of Jerusalem, Department of Developmental Biology and Cancer Research, Jerusalem, Israel

Neutrophils are critical players in the innate immune response, being among the first cells to detect pathogens and respond to infection. Upon detection, neutrophils eliminate pathogens through phagocytosis, degranulation, and decondense chromatin to release Neutrophil Extracellular Traps (NETs)—web-like structures composed of chromatin, histones, and antimicrobial proteins that trap and neutralize pathogens. NET formation is an essential mechanism for controlling infections, however this process can become maladaptive, leading to dysregulated immune conditions, such as thrombosis and sepsis. Studying NET formation in isolated neutrophils is important for understanding underlying mechanisms but it does not account for the complexity of immune interactions that occur in whole blood, limiting our understanding of neutrophil responses. For the first time we describe chromatin changes that occur during the early stages of NET formation that occur within whole blood. This study compares chromatin accessibility dynamics during NET formation by stimulating primary human neutrophils with phorbol 12-myristate 13-acetate (PMA) following isolation or in the context of whole blood. An ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing) time-course was conducted to profile chromatin changes preceding NET formation. We compared whole blood stimulation to PMA-stimulated isolated neutrophils to assess the impact of immune cell crosstalk. We identified distinct chromatin accessibility patterns across isolated-only, whole blood-only, and overlapping regions, and found that whole blood provides a more diverse range of both open and closed chromatin states in response to PMA induced NET formation. This complexity was reflected in the transcription factor motifs enriched in the differentially accessible regions, where whole blood and overlapping regions showed greater association with NET-related and inflammatory transcription factors, while isolated neutrophils showed fewer relevant motifs. Additionally, pathway analysis indicated that whole blood responses involved more robust activation of immune-specific pathways, such as interleukin and cytokine signaling, compared to isolated neutrophils. These findings emphasize that chromatin decondensation during NET formation is influenced by interactions within whole blood, underscoring the importance of studying NET formation in a clinically relevant environment to identify molecular biomarkers and develop NET inhibitors.

Immunological and epigenetic changes induced by high-radiation spacewalks

Christopher E. Mason[1,2,3,4], Jiwoon Park[1,2], JangKeun Kim[1,2], Eliah Overbey[5], Jeremy Wain Hirschberg[1,2], Yui Ci[6], Martin Shelton[6], Shanshan He[6], Jacqueline Proszynski[1,2], Krista Ryon[1,2], Paul Collier[1,2], Cem Meydan[1,2], Jasmine Plummer[7], Luciano G. Martelotto[8]

1 Cornell University, Weill Cornell Medicine, Department of Physiology and Biophysics, New York, NY
2 Cornell University, Weill Cornell Medicine, The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, New York, NY
3 Cornell University, Weill Cornell Medicine, The WorldQuant Initiative Quantitative Prediction, New York, NY
4 Cornell University, Weill Cornell Medicine, The Feil Family Brain & Mind Research Institute, New York, NY
5 University of Austin (UATX), Austin, TX
6 Bruker Inc., Middlesex, MA
7 St. Jude's Children Hospital, Memphis, TN
8 The University of Adelaide, Adelaide, South Australia

Spaceflight presents a unique environment with profound physiological effects, providing an opportunity to study human adaptation to extreme conditions and to aid in preparations for future missions. To better understand the spaceflight environment, we have analyzed more than 60 astronaut samples collected as part of the SOMA (Space Omics and Medical Atlas), including the high-radiation Polaris Dawn (PD) 2024 mission into the Van Allen belts. For the PD samples, we utilized several new single-cell, spatial, and epigenetic technologies to create the first-ever map of changes associated with a high-elevation, high-radiation environment and spacewalks at 1400km. Specifically, we performed paired spatial multi-omics assays from peripheral blood mononuclear cells (PBMCs) from multiple missions (CosMx spatial WTx and STAMP), coincident epitope mapping (PixelGen), deep genome sequencing for rare variants (Ultima and Illumina), and targeted and overall protein changes (Olink and SOMALogic).

Our findings reveal distinct transcriptomic and proteomic alterations during spaceflight and spacewalking, including upregulation of stress-response pathways, immune modulation, DNA damage, and hematopoietic activity. Changes in T cells and monocytes indicate altered activation, immune surveillance, and inflammatory responses. Hematopoietic stem cells (HSCs) showed adaptive signaling changes linked to stemness and telomere length. Still, even single-cell sequencing analyses were insufficient to identify these rare cells and perform deep profiling. However, whole-transcriptome spatial omics sequencing addressed this issue by enabling higher throughput profiling of more cells (>10M) per sample and characterizing subcellular transcript distributions, revealing T-cells as more responsive to spaceflight and mutations than B-cells, dendritic cells, or other immune cell populations.

This study provides a detailed, multi-omics perspective on temporal molecular changes in astronauts, highlighting pathways that may contribute to immune dysregulation or adaptation. These findings enhance our understanding of human health in space and inform personalized medical strategies for astronauts during long-duration missions, and can guide studies on new biomarkers for tracking the efficacy of countermeasures, as well as to prepare humans for exploration-class missions to the Moon and Mars.

Identification of novel biological functions through analysis of Y chromosome genome structure based on the haplotypes

Yoko Kuroki[1], Yutaka Suzuki[2]

1 National Center for Child Health and Development, Department of Genome Medicine, Setagaya, Tokyo, Japan
2 University of Tokyo, Life Science Data Analysis Center, Kashiwa, Chiba, Japan

The Y chromosome is a haploid genome unique to males with no genes essential for life. It is transmitted to the next generation without being repaired by recombination, even if an extensive genomic structural alteration occurs. On the other hand, the Y chromosome genome is basically a region transmitted only from father to son, reflecting a male-specific inheritance between generations. The Y chromosome exhibits genomic structural differences among different ethnic groups and individuals. The Y chromosome was previously thought to affect only male-specific phenotypes. However, recent studies have revealed associations between the Y chromosomes and phenotypes common to both males and females, such as certain types of cancer and neuropsychiatric disorders.

We aim to reveal novel genomic functions of the Y chromosome related to the biological functions to maintain a healthy condition or susceptibility to several diseases. We will conduct a comparative analysis of the genomic structure of the Y chromosome in patients with diseases involving the sex chromosome, such as azoospermia and disorders of sex differentiation, and the genomic structure of the Y chromosome in healthy individuals. We build haplotype-specific Y chromosome reference sequences using a Y chromosome-specific genomic library of Japanese male origin that we have constructed ourselves. We will use these reference sequences to distinguish the genome structure and the number of gene copies for Y-linked genes in each haplotype. In addition, we will perform Y chromosome haplotype analysis of patients with azoospermia, disorders of sex differentiation, and healthy controls. Then, we will identify the genetic regions involved in disease onset by comparing the haplotype frequencies of the two groups. Our project is ongoing, and we will report on the latest progress at this meeting.

Beyond Risk: Genetic factors of type 2 diabetic nephropathy resilience in Black/African Americans

Tennille Leak-Johnson[1], Hannah Luk[2], Kenisha Webb[1], Ariel Williams[3]

1 Morehouse School of Medicine, Atlanta, GA, USA
2 University of Texas-Medical Branch John Sealy School of Medicine, Galveston, Texas, USA
3 National Institute of Health/ National Human Genome Research Institute, Rockville, MD, USA

Background: The engulfment and cell motility 1 (ELMO1) gene encodes a soluble cytoplasmic protein crucial for cytoskeletal rearrangement, phagocytosis of apoptotic cells, and cell motility. Variants in ELMO1 have previously been associated with protection in Black/African Americans (B/AA) with type 2 diabetes (T2D) with end-stage renal disease (ESRD). The exact mechanisms by which ELMO1 contributes to T2DN protection in African-derived populations have yet to be fully understood.

Methods: We utilized the All of Us Research Program (AoURP) to replicate four well-established variants in intron 8 of the ELMO1 gene and conducted preliminary analysis of 14 high-frequency African ancestral nonsynonymous SNPs in genes that promote cell motility and phagocytosis signaling pathways in candidate genes BCAR1, ELMO3, DOCK2, and ELMO2. Associations with T2DN status and quantitative traits HgA1c and GFR were evaluated in 2,245 B/AA normoglycemic T2D controls and 834 T2DN cases.

Results: Single-SNP allelic association analyses adjusted for age, sex at birth, BMI, and HgA1c confirmed associations with rs1345365 (OR = 0.67; P = 0.014) and rs10951509 (OR = 0.73; P = 0.034), both of which are in strong linkage disequilibrium (r2 = 0.97) within an enhancer region. Nominal associations were observed with DOCK2 and T2DN protection (OR=0.65; P = 0.036) after adjusting for age, sex at birth, BMI, HgA1c, and GFR. The most significant associations observed were between increased GFR and rs2058730, also located in the intronic region of ELMO1 (age, sex at birth, BMI, and HgA1c-adjusted P = 0.003) and DOCK2 SNP with decreased HgA1C (age, sex at birth, BMI, and GFR-adjusted P = 0.001).

Conclusion: These findings suggest that genetic variants in ELMO1 and DOCK2 may protect B/AA individuals with T2D from nephropathy progression, or complex interactions with psychosocial factors may influence resilience, with risk and resilience variants acting orthogonally within the ELMO1 pathway.

DNA methylation landscape of nematode-infected cattle: infection-related patterns and immune response pathways

George E. Liu[1*], Clarissa Boschiero[2,3], Ethiopia Beshah[2], Xiaoping Zhu[3], Wenbin Tuo[2*]

1 United States Department of Agriculture (USDA), Agricultural Research Service (ARS), Animal Genomics and Improvement Laboratory, Beltsville Agricultural Research Center (BARC), Beltsville, MD
2 United States Department of Agriculture (USDA), Agricultural Research Service (ARS), Beltsville Agricultural Research Center (BARC), Animal Parasitic Diseases Laboratory, Beltsville, MD
3 University of Maryland, Department of Veterinary Medicine, College Park, MD
*Corresponding authors

DNA methylation (DNAm) regulates gene expression and genomic imprinting, influencing traits like growth, reproduction, and disease resistance. While DNAm patterns in specific cattle tissues have been studied, the effects of gastrointestinal (GI) nematode infection on DNAm remain largely unknown. Helminth-free Holstein steers (n=5) were trickle-infected orally with O. ostertagi stage 3 larvae in tap water (1,000 larvae/day) for 4 weeks, 5 days per week. Control animals (n=5) received tap water only. All animals were euthanized 30 days after the first infection and tissues were collected and stored at -80°C until used. We conducted epigenome-wide profiling using the mammalian methylation array (HorvathMammalMethylChip40) and explored the impact of GI nematode infection on the methylation patterns in four different cattle tissues including mucosal tissues from duodenum (DUO), pyloric abomasum (PYL), and fundic abomasum (FUN) and abomasal draining lymph nodes (dLN) in infected and uninfected animals. The analysis covered 31,107 cattle CpGs of 5,082 genes derived from 37 samples (19 infected and 18 noninfected tissues) and revealed infection-related, tissue-specific, differential methylation patterns. A total of 389 shared and 2,770 tissue-specific differentially methylated positions (DMPs, 5% Δ□) were identified in dLN and FUN, particularly in genes associated with immune responses. The shared DMPs were found in 263 genes where those relating to immune-response pathways are abundant (BCL11A, IFNG, IL20RB, IKZF1, LRFN5, NTRK2, PSMD7, SATB1, SMAD7, TFEC, TLE1, TLE3, TP63, ZEB2). Furthermore, 282, 244, 52, and 24 differentially methylated regions (DMRs, 10% Δ□) were observed in dLN, FUN, PYL, and DUO, respectively. There were more hypomethylated DMRs detected in dLN and FUN, while more hypermethylated DMRs were found in PYL and DUO. Among the genes carrying DMPs and DMRs, immune-related genes including BCL11A, EPHB2, EGR3, FKBP5, FOSB, IL17B, IL20RB, IRX5, NTRK3, OIT3, RAB2A, RARA, SATB1, SATB2, SMAD4, TLE1/4, TLX3, and ZEB2 were revealed. We further uncovered enriched pathways related to immune functions in infected animals, indicating a link between DNAm changes and GI nematode infection. These pathways included WNT/□-catenin signaling, Th1 and Th2 activation, RAR activation, NGF-stimulated transcription, IL-15 production, and IL-4, 7, 12, and 13 signaling pathways. The data may implicate a crucial role of DNAm in regulating the nature and/or strength of host immunity to the infection, which will facilitate further understanding of immune evasion, a survival strategy commonly used by parasitic nematodes. Integration with RNA sequencing data additionally revealed genes related to immune functions, some of which are transcription factors. These findings contribute to a deeper understanding of the epigenetic regulatory landscape in cattle during GI nematode infection, highlighting potential targets for further investigation.

Keywords: cattle, O. ostertagi, immune response, DNA methylation, duodenum, lymph nodes, abomasum.

Whole transcriptome (WTX) spatial atlas of cellular senescence in the human post-menopausal ovary

Nicolas Martin[1], Pooja Raj Devrukhkar[2-3], Hannes M. Campo[2-3], Josef Byrne[1,4], Bikem Soygur Kaya[1], Mark Watson[1], Wei Yang[5], Haiyan Zhai[5], Joseph Beechem[5], Shanshan He[5], Birgit Schilling[1], Francesca E. Duncan[1-3], Simon Melov[1,4]

1 Buck Institute for Research on Aging, Novato, CA
2 Northwestern University, Feinberg School of Medicine, Department of Obstetrics and Gynecology, Chicago, IL
3 Northwestern University, Feinberg School of Medicine, Department of Pathology, Chicago, IL
4 University of Southern California, Leonard Davis School of Gerontology, Los Angeles, CA
5 Bruker Spatial Biology, Inc., Seattle, WA

The human ovary is among the first organs to experience early functional decline, impacting reproductive longevity and overall systemic health. The underlying causative mechanisms that drive functional decline have not been fully elucidated. We hypothesize that cellular senescence significantly contributes to ovarian aging, potentially promoting fibrosis and inflammation. However, characterizing senescent cells in the aging ovary is particularly challenging due to the heterogeneity of the tissue and the high level of cell density relative to other tissues.

By utilizing our snucRNA-seq human post-menopausal ovarian atlas, which comprises over 500,000 nuclei from 28 post-menopausal subjects, we developed a method to detect senescent cells directly within native ovarian tissue. This approach involves whole transcriptome profiling at the single-cell level, allowing for the spatial identification of senescent cells based on their transcriptomic signatures. Spatial profiling was conducted using tissue microarrays containing cortical ovarian tissue from 40 post-menopausal subjects.

With over 18,000 genes probed in situ, we identified a substantial population of stromal cells expressing canonical markers of cellular senescence, including LAMB1, TIMP3, and SERPINE1. Single-cell spatial characterization revealed at least 16 distinct cell types within cortical ovarian tissue, with several sub-clusters showing upregulation of previously reported senescence markers, such as CDKN1A, GADD45B, and ADAMTS4.

The ability to map and characterize senescent cells via in situ whole transcriptome profiling will allow us to further understand the role of senescent cells in the post-menopausal human ovary.

Deciphering the effects of RNA N6-methyladenosine on gene expression in livestock

Stephanie D. McKay[1], Shangqian Xie[2], Brenda M. Murdoch[2]

1 University of Missouri, Division of Animal Sciences, Columbia, MO
2 University of Idaho, Department of Animal, Veterinary and Food Sciences, Moscow, ID

Determining the extent of epigenetic effects upon phenotypic variation involves characterization of both the epigenome and the epitranscriptome. Both DNA 5-methylcytosine (5mC) and RNA N6-methyladenosine (m6A) are common epigenetic modifications that play a role in transcription and post-transcriptional regulation, respectively. Recent evidence suggests the existence of a potential mechanism of coordinated transcriptional (5mC) and post-transcriptional (m6A) regulation in various biological processes, further emphasizing the need for epitranscriptomic annotation. Therefore, with the aim of elucidating the effects of DNA and RNA methylation on gene expression, we have performed Whole Genome Bisulfite Sequencing (WGBS), Methylated RNA Immunoprecipitation Sequencing (MeRIP-Seq) and RNA-Seq in three tissues from four individuals of two different animal species. Initially, DNA 5mC density was quantified for each sample and results demonstrated that the 5mC density was similar in three tissues (caruncle, spleen, and mammary gland) of sheep (2.04%, 2.62% and 2.53%) and cattle (2.52%, 2.77% and 2.61%). RNA m6A modifications identified from MeRIP data in the same three tissues yielded a total of 10,295, 19,874 and 5,629 peaks were identified in sheep, and 14,377, 18,698 and 14,339 peaks were identified in cattle. Subsequently, comprehensive genome-wide self-interaction and across-interaction networks of DNA methylation, RNA methylation and gene expression were constructed to assess the relationships of 5mC, m6A and gene expression. The results from both species consistently indicated that RNA methylation exerts a more substantial effect on gene expression than DNA methylation, showing a significant positive correlation. Interestingly, methylation occurring within gene bodies had a more pronounced impact on gene expression compared to that found in promoter regions. This work offers new perspectives and opportunities to advance epigenomic and epitranscriptomics research by deepening our understanding of genome-wide dynamics and the complex associations among DNA methylation, RNA methylation and gene expression.

Transcriptomic characterization of host response to Nipah virus infection

Monika Mehta[1], Lirong Peng[1], Shawn Hirsch[1], David Drawbaugh[1], Saurabh Dixit[1], Sanae Lembirik[1], Erin Kollins[1], Russ Byrum[1], Yu Cong[1], Michael R. Holbrook[1]

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

Nipah virus (NiV), a highly pathogenic paramyxovirus (genus Henipavirus), is an emerging zoonotic pathogen responsible for outbreaks of acute respiratory and neurological disease. The high fatality rates associated with NiV infections, coupled with the lack of effective medical countermeasures approved for human use, have prompted the World Health Organization to designate Nipah virus as a priority pathogen for research and development.

The Integrated Research Facility at Fort Detrick uses a multidisciplinary approach to study high-consequence diseases caused by novel, emerging, and highly virulent biosafety level 4 (BSL-4) viruses. Previously, we developed animal models of NiV infection and evaluated host responses to exposure to the Bangladesh and Malaysia isolates of the virus (Lee et al., 2020; Hammoud et al., 2018). Recently, we extended those studies to include in-depth genomic analyses to examine the host pulmonary and neurological changes. Specifically, we generated transcriptomic profiles of lung and brain tissue samples from NiV-exposed green monkeys (Chlorocebus sabaeus) and compared the transcriptomic profiles of animals that survived with those that succumbed to disease. Additionally, we compared host responses to the two isolates and correlated the transcriptomic data with the pathological changes observed in the animals. Furthermore, we analyzed the immune cell populations in blood samples using single-cell RNA sequencing and examined the dominant T-cell and B-cell receptor clonotypes in animals that survived or succumbed to disease caused by exposure to the two NiV isolates. Virus sequence analyses revealed changes in the virus genomes over the disease course.

This presentation will describe the conclusions from these studies and discuss some of the unique challenges associated with genomic analyses of high containment pathogens.

Contribution of rare structural variants and tandem repeats to autism from long-read whole genomes

Milad Mortazavi[1], James Guevara[1], Joshua Diaz[1], Sergey Batalov[6], Matthew Bainbridge[6,8], Abraham Palmer[1,4], Aaron Besterman[1,6,7], Melissa Gymrek[2,3], Jonathan Sebat[1,4,5]

1 University of California San Diego, Department of Psychiatry, La Jolla, CA
2 University of California San Diego, Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA
3 University of California San Diego, Department of Medicine, La Jolla, CA
4 University of California San Diego, Institute for Genomic Medicine, La Jolla, CA
5 University of California San Diego, Department of Cellular and Molecular Medicine and Pediatrics, La Jolla, CA
6 Rady Children's Institute for Genomic Medicine, San Diego, CA
7 Rady Children's Hospital San Diego, San Diego, CA
8 Codified Genomics LLC, Houston, TX

Long read whole genome sequencing (WGS) was performed on 243 subjects from 63 autism spectrum disorder (ASD) families recruited to the REACH study at Rady Children's Hospital and UC San Diego, including 101 genomes sequenced to 8X with Oxford Nanopore (ONT) and 142 genomes sequenced to 3-5X with Pacific Biosciences HiFi sequencing. SV and TR calling of all genomes was performed using a common pipeline consisting of alignment (minimap2), SV detection (Sniffles2, lumpy), SV genotyping (SnoopSV) and TR genotyping (LongTR). We investigated genetic association of rare SVs and TRs in genes stratified by functional element (exon or intron) and functional constraint (pLI > 0.9), testing variant burden in families by conditional logistic regression, controlling for sex, ancestry principal components, sequencing coverage and genome-wide SV/TR burden. Nominally significant associations ($P < 0.05$) were detected for rare exonic SVs and TRs in constrained genes, and no association was found for intronic and intergenic variants. Rare exonic SVs and TRs from long read WGS explained a combined 3.5% of the variance in case status which was comparable to the variance explained by loss-of-function (LoF) SNVs (4.9%) and polygenic scores (2.3%) from Illumina WGS of the same cohort (PMID: 35654974). In addition to the identification of novel risk variants, sequence-level resolution of SVs from long read WGS provided information necessary for clinical interpretation. We present the case of a 17 year old inpatient with a diagnosis of autism carrying a duplication of 9p24.2. PacBio HiFi WGS of the family revealed the duplication to be part of a complex rearrangement including partial deletion and loss of function of RFX3, a gene in which LoF variants cause intellectual disability (PMID: 33658631). Analysis of combined short read and long read genomes confirmed that RFX3 mutation was transmitted from a parent who was mosaic for the complex SV. Our results demonstrate that analysis of SVs and TRs by long read WGS captures substantial variance in ASD not explained by rare and common SNVs and functional information not evident by short read sequencing.

Decoding the genome of bighorn sheep: implications for species survival

Brenda M. Murdoch[1*], Temitayo Olagunju[1*], Yana Safonova[2], Stephanie D. McKay[3], Benjamin D. Rosen[4‡], Timothy P.L. Smith[5‡], and the Sheep Pangenome Consortium

1 University of Idaho, Animal, Veterinary and Food Sciences (AVFS), Moscow, ID
2 Pennsylvania State University, Computer Science and Engineering Department, University Park, PA
3 University of Missouri, College of Agriculture, Food and Natural Resources, Columbia, MO
4 United States Department of Agriculture (USDA), Agricultural Research Service (ARS), Animal Genomics and Improvement Laboratory, Beltsville, MD
5 United States Department of Agriculture (USDA), Agricultural Research Service (ARS), U.S. Meat Animal Research Center, Clay Center, NE
*Co-first authors
‡Co-corresponding authors

Bighorn sheep (Ovis canadensis) populations in North America have historically numbered between 150,000 and 200,000 but faced dramatic declines by 1900 due to habitat destruction, overhunting, and disease. Although North American wild sheep diverged from domestic sheep approximately 5.3 million years ago, they still interact when given the opportunity, leading to increased disease vulnerability. Current challenges include high morbidity and mortality rates from pneumonia-associated diseases presumably transmitted from domestic sheep, goats and other wildlife. While domestic sheep can suffer from pneumonia it is typically not as severe or fatal as in bighorn sheep. We developed telomere-to-telomere (T2T) genome assemblies and a super-pangenome to enhance genomic resources that capture the genetic diversity of sheep, including bighorn sheep. By advancing genome assemblies to T2T and employing a non-linear graph to represent alternative haplotypes, we aim to improve our ability to evaluate structural variants and indels alongside single nucleotide polymorphisms. Our research catalogs variants across 23 breeds and four subspecies of sheep, examining the evolutionary roles of these variants in response to natural and artificial selection within specific environments. Interestingly, adaptive immune loci, such as immunoglobins and T-cell receptor loci, are extremely diverse within and across sheep and bighorn sheep. Understanding these differences in genetic variation in immune loci of domestic and bighorn sheep will empower scientists to better understand why and how physiological and immune responses vary. The resulting T2T assemblies and super-pangenome will serve as invaluable tools for uncovering the mechanisms underlying biological variation important for sheep conservation and as a global resource of food and fiber.

Keywords: super-pangenome, T2T, Bighorn, ovine,

Comprehensive genetic screening for SMA in the Dutch newborn screening program: enhancing diagnosis and family counseling

Mirjam S. de Pagter[1*], Maria M. Zwartkruis[*], Marije Koopmans[1], Mirjan Albring[1], Inge Cuppen[2], Els Voorhoeve[3], Nienke E. Verbeek[1], Gijs van Haaften[1], W. Ludo van der Pol[2], Ewout J.N. Groen[2,^], Marcel R. Nelen[1,^]

1 University Medical Center (UMC) Utrecht, Department of Genetics, Division Laboratories, Pharmacy and Biomedical Genetics, Utrecht, the Netherlands
2 University Medical Center (UMC) Utrecht, UMC Utrecht Brain Center, Department of Neurology and Neurosurgery, Utrecht, the Netherlands
3 National Institute for Public Health and the Environment (RIVM), Centre for Health Protection, Bilthoven, The Netherlands
*/^These authors contributed equally.

Spinal muscular atrophy (SMA) is an autosomal recessive disorder marked by muscle weakness and atrophy. If untreated, the condition progresses rapidly, leading to respiratory failure and death before the age of two in the majority of patients affected by its most common, and severe, form. SMA is primarily caused by a homozygous deletion of (part of) the SMN1 gene. In June 2022, SMA was added to the Dutch newborn screening program, focusing on detecting this specific homozygous SMN1 gene deletion. Upon a positive newborn screening for SMA, genetic testing is conducted in our laboratory using a fresh peripheral blood sample. This test has an average turnaround time of two to three workdays, confirming the initial screening results and determining the SMN2 copy number. SMN2 copy number is crucial for predicting disease severity and determining treatment options in (presymptomatic) newborns. In its first two years, the screening program identified 31 newborns with SMA, which is consistent with the expected numbers. No false positives or negatives were detected during this period. Thus, genetic newborn screening for SMA far exceeds the accuracy of tests used for other disorders in the newborn screening program.

Following newborn screening, most parents of a child with SMA undergo carrier testing. Although the vast majority of parents are identified as carriers, over 11% of those tested do not appear to be carriers. A major challenge in this context is determining whether a de novo deletion has occurred or if the parent is a silent carrier (i.e., possessing two copies of SMN1 on one allele and none on the other). This distinction is critical as it directly influences the recurrence risk assessment for future pregnancies. To address this challenge, we are developing an advanced methodology utilizing a combination of established techniques and state-of-the-art long read sequencing. Our goal is to enhance the accuracy of carrier detection and differentiate between inherited and de novo deletions. This initiative is expected to enhance the reliability of genetic counseling for families, ensuring that parents and family members receive accurate information about the risk of recurrence in future pregnancies.

Overall, the inclusion of SMA in the Dutch newborn screening program marks a significant advancement in pediatric healthcare, also demonstrating the accuracy and efficacy of genetic testing in newborn screening. By integrating advanced genetic testing and established methods to distinguish between silent carriers and de novo deletions, we aim to provide comprehensive care and informed guidance to affected families.

The Spatial Atlas of Human Anatomy (SAHA): comprehensive multi-omics spatial mapping of human organ systems to improve precision medicine

Jiwoon Park[1], Roberto De Gregorio[1], Brian Robinson[1], Erika Hissong[1], Sanjay Patel[1], Rohit Chandwani[1], Fabio Socciarelli[1], Patrick Danaher[2], Evelyn Metzger[2], Andrew Walters[3], Sajad Darabi[3], Harry Clifford[3], Steven Kothen-Hill[3], Massimo Loda[1], Olivier Elemento[1], George Vacek[3], Joseph Beechem[2], Alicia Alonso[1], Shauna Lee Houlihan[1], Robert E. Schwartz[1], Christopher E. Mason[1,*]

1 Cornell University, Weill Cornell Medicine, New York, NY
2 Bruker Spatial Biology, Seattle, WA
3 NVIDIA Corporation, Santa Clara, CA

The Spatial Atlas of Human Anatomy (SAHA) maps the cellular and molecular landscapes of 30 human organs from healthy adults at an unprecedented spatial resolution (50 nm). This large-scale, collaborative project sets a new benchmark in spatial precision medicine, establishing best practices for experimental design, sample processing, and data standards. Capturing variability across genders and ancestries, the dataset serves as a valuable reference for understanding human biology at an unmatched level. Since its initial presentation at the 2022 AGBT meeting, SAHA has achieved significant advancements, including the completion of organ maps and the standardization of spatial multi-omics techniques. This year, we (1) report detailed maps for the digestive and immune systems, (2) showcase the highest subcellular resolution maps of cell types, metabolic capacities, and ligand-receptor interactions, and (3) preview the next phase of multi-omics integration and precision oncology applications.

SAHA currently encompasses immune (bone marrow, lymph node, PBMC) and gastrointestinal (ileum, appendix, colon, liver, pancreas, stomach) tissues, as well as diseased samples, including Crohn's disease and cancers. By integrating multiple cutting-edge spatial platforms (CosMx™ Spatial Molecular Imager, Xenium, RNAscope, and GeoMx® Digital Spatial Profiler) with histopathology and sequencing, we have mapped over 10 million cells from 100+ individuals, providing insights into cellular diversity at both tissue and subcellular scales. Distinct immune and epithelial landscapes in each organ highlight functional specialization and microbial interactions, such as ephrin (EFN) signaling patterns. A comparison of pancreatic cancer samples with normal pancreas datasets reveals a distinct T-cell-rich tumor microenvironment linked to aggressive genotypes, providing insights into patient heterogeneity. These findings lay the foundation for next-generation diagnosis tools and personalized therapies.

To maximize the accessibility and impact of spatial data, we developed spatialGPT, a generative AI model trained on SAHA data that predicts cell types, infers cross-modal information, and identifies spatial patterns linked to disease progression. Alongside a web-based data exploration portal, SAHA provides the broader scientific community with unprecedented access to high-resolution data. We plan to expand SAHA to additional organ systems including reproductive systems and develop comprehensive roadmaps for integrating spatial omics with clinical applications, setting a new paradigm for precision medicine.

A tale of two epidemics: a spatial transcriptomic study of fentanyl effects on type II diabetes

Megan Pater[1], Grant R. Kolar[1], Joshua D. Stafford[1], Willis K. Samson[1], Gina L.C. Yosten[1]

1 Saint Louis University School of Medicine, Department of Pharmacology and Physiology, Saint Louis, MO

Background: Synthetic opioids, such as fentanyl, are the leading cause of drug overdose. It is understood that opioids have hyperglycemic effects. However, the potential effects of fentanyl abuse on pancreatic islet cell transcriptomic profiles have yet to be fully described, even though the delta opioid receptor has been identified in human islets. Spatial molecular imaging (SMI) offers a novel opportunity to determine at the cellular and molecular level these potential metabolic alterations.

Methods: Pancreatic tissues from 187 patients were analyzed utilizing the Nanostring CosMx SMI [6000-plex]. Data was analyzed using R 4.0, Seurat v5 and sf packages. Electronic health records (EHR) were used to stratify patients based on fentanyl screening and A1c levels. The spatial transcriptomic profiles of fentanyl positive patients were compared to non-fentanyl users and diabetic groups. Identifying genes for endocrine cells were also compared across these groups.

Results: We observed an upregulation in hypoxia-related genes [HIF1A and FLT1] in pancreatic islets of fentanyl positive patients, compared to fentanyl negative patients. Transcriptional profiles of fentanyl positive patients positively correlated those observed in prediabetic patients, suggesting the possibility that fentanyl usage either directly or indirectly compromises islet cell function.

Conclusions: We believe that hypoxic damage in the endocrine portion of the pancreas leads to the development of hyperinsulinemia. Undamaged beta cells increase insulin expression to compensate for beta cell loss, which may lead to further oxidative damage and loss of beta cell mass. Our lab has acquired pancreas tissues from over 800 cases considered for transplantation, all accompanied with EHRs, which allows us to query how demographics and potential substance abuse may contribute to endocrine disorders often observed across these populations. Future studies will examine potential alterations in other protective pathways in endocrine cells of patients with history of fentanyl use.

Brain vascular changes in aging and Alzheimer's: insights from VINE-seq and spatial transcriptomics

Alex J. Pennacchio[1], Madigan Reid[1,*], Tammie Tam[2,*], Bella Ding[1], Brittany Davidson[2], Jessica Tsui[2], Amanda Apolonio[1], Hao Liu[1], Shih-Hsiu J. Wang[3], Alexis Combes[2,#], Andrew C. Yang[1,#]

1 Gladstone Institutes, Gladstone Institute of Neurological Disease, San Francisco, CA
2 University of San Francisco, Department of Pathology, San Francisco, CA
3 Duke University Medical Center, Department of Pathology, Durham, NC
*Contributed equally
#These authors jointly supervised this work.

The blood-brain barrier (BBB) is a selectively permeable protective structure that regulates the transport of necessary molecules while protecting the brain from harmful substances. Formed by a specialized community of vascular and immune cells, the BBB is essential for maintaining healthy brain function. While some BBB degradation occurs with normal aging, this breakdown is exacerbated in individuals with Alzheimer's disease (AD), as seen through increased leakage of blood components in histopathological and clinical imaging studies. Brain vascular dysfunction is now recognized as one of the earliest pathological changes associated with late-onset AD. For example, disruptions in cerebral blood flow and damage to the BBB can precede clinical symptoms by decades. However, molecular changes to the BBB in aging and disease are not well characterized, in part due to technical shortcomings of traditional scRNA-seq workflows in detecting functionally important vascular cell types.

The recent development of vessel isolation and nuclei extraction for sequencing (VINE-seq), enables large-scale capture of vascular cell types, which have been under-represented in previous AD scRNA-seq studies. We generate scRNA-seq data of paired cortical gray and adjacent white matter post-mortem tissue from a cohort of 107 individuals. Our cohort includes healthy individuals spanning all decades of adult life (ages 21 to 89), individuals with AD, and aged resilient donors- those with AD-like brain pathology but no cognitive symptoms.

Using VINE-seq, we capture over 1,000,000 cells, including more than 150,000 vascular cells, which we utilize for downstream analyses. We apply computational factorization methods to identify programs of genes that show coordinated expression changes across multiple cell types during aging and disease. Additionally, leveraging the wide age range of our cohort, we perform gene expression trajectory analysis to establish a baseline for healthy aging at the cell-type level, and examine divergences from these trajectories in diseased individuals. Finally, we employ 10x Xenium spatial transcriptomics to understand how vascular cell types change expression in situ locally around AD pathologies and corresponding changes in immune cell distribution and activation state. Our findings offer new insights into the vascular and immune contributions to aging and AD at single-cell resolution.

DNA methylation dynamics of mammalian neuronal maturation

Lauren E. Rylaarsdam[1], Ruth V. Nichols[1], Brendan L. O'Conell[1], Zohar Shipony[2], Gail Mandel[3], Andrew C. Adey[1]

1 Oregon Health & Science University, Molecular & Medical Genetics, Portland, OR
2 Ultima Genomics, Fremont, CA
3 Oregon Health & Science University, Vollum Institute, Portland, OR

Cell specification and cortical development depend on precise gene expression, often regulated by epigenetic changes such as DNA methylation present at cytosine residues (DNAm). While DNMT1 directs methylation at cytosines followed by guanine (mCG) in most tissues, neurons uniquely accumulate methylation at cytosines followed by non-guanine bases (mCH), deposited by DNMT3A during early postnatal development. This process coincides with key neurodevelopmental events, including peak synaptogenesis; however, the dynamics of mCH accumulation and mCG remodeling have yet to be profiled with cell type granularity.

We recently developed an atlas-scale single-cell DNA methylation technology, sciMETv3, which enables the production of hundreds of thousands of single-cell methylomes in a single experiment. We demonstrate that sciMETv3 is compatible with both Illumina and Ultima sequencing platforms, allowing the utilization of recent advances in high-throughput sequencing. We also developed the ability to profile both chromatin accessibility (ATAC) and DNAm from the same cells, enabling dataset-bridging and the exploration of cross-property interactions. To enable the analysis of large-scale single-cell DNAm datasets we developed an analysis platform, Amethyst, that enables any user to perform a comprehensive analysis and interaction of such studies.

Here, we apply sciMETv3 and Amethyst to profile neuronal methylomes across five regions of the mouse brain during the key window of neurodevelopment when noncanonical DNAm is deposited. Across both regions and cell types, we observe significant differences in the rate, levels and gene specificity of mCH accumulation. Using our atlas, we produced ordered trajectories of the dynamics of DNA methylation across the breadth of neuronal cell types, framing the role and impact of noncanonical methylation as well as distinct, focal mCG changes with direct gene pathway implications. These results during the key period of synaptogenesis may provide key insight into the timing and nature of epigenetic disruptions in neurodevelopmental disorders.

Evaluation of the Ultima Genomics UG100 sequencer for affordable metagenomic diagnosis of meningitis and encephalitis

Ryan C. Shean[1], David Bogumil[2], Alix Cruse[2], Samm Hernadez[2], Nika Iremadze[2], Sarah Pollock[2], Doron Lipson[2], Benjamin T. Bradley[1]

1 University of Utah and Associated Regional and University Pathologists (ARUP) Laboratories, Salt Lake City, UT
2 Ultima Genomics, Fremont, CA

Introduction:
Metagenomic next-generation sequencing (mNGS) is an advanced diagnostic method that uses high-throughput technology to simultaneously sequence millions of DNA fragments. This technique is particularly valuable in diagnosing meningitis and encephalitis (M/E), as it can identify virtually any infectious agent, including bacteria, viruses, fungi, and parasites, directly from cerebrospinal fluid (CSF) in a single assay. However, the broad adoption of mNGS for diagnosing infectious meningitis and encephalitis has been limited by high costs and poor sensitivity. Here, we present a proof-of-concept study evaluating the performance of the Ultima Genomics sequencing platform for low-cost mNGS of clinical CSF samples.

Methods:
Libraries were prepared using 10-20 ng DNA isolated from CSF samples following the P00014_UG_NEB PCR-free WGS Library Prep_RevD protocol. To accommodate low DNA amounts, adapter input was diluted 1:10 before ligation, and libraries were PCR-amplified for 15 cycles. Sequencing was performed on the UG100 Sequencer, generating single-end 300 bp reads. A human genome control (HG002) was included for quality assurance. Reads unaligned to HG38 were assembled de novo (Megahit), aligned to a database of the expected pathogen genomes and taxonomically classified using Kraken2.

Results:
A total of 20 CSF samples were sequenced on a single $2000 wafer, including 7 CSF samples from clinically confirmed cases of M/E. Ultima sequencing identified the correct pathogen in 100% (11/11) of contrived samples and 86% (8/9) of clinical specimens. Several interesting observations were made in this proof-of-concept study. First, a near-complete genome of Haemophilus influenzae was recovered from one clinical specimen; traditional PCR tests for diagnosing H. influenzae meningitis do not provide this level of resolution. Second, reads (n=13,409) mapping to Taenia solium were identified in a patient with serology results indicating infection with this pathogen. Currently, no FDA-approved molecular test for T. solium exists. Ultima also detected reads for Borrelia burgdorferi, the causative agent of Lyme disease, in a clinical sample near the limit of detection (CT 39.4) by a lab-developed nucleic acid amplification test.

Conclusion:
Metagenomic sequencing on the UG100 platform provides accurate pathogen detection and a low per-sample cost.

Spatial characterization of the immune response in a long-term pig-to-human kidney xenotransplantation

Claire Williams[1], Valentin Goutaudier[2], Erwan Morgand[2], Fariza Mezine[2], Jeffrey Stern[3], Karen Khalil[3], Adam Griesemer[3], Sébastien Giraud[4], Thierry Hauet[4], Massimo Mangiola[3], Wei Yang[1], Christina Bailey[1], Haiyan Zhai[1], Joseph Beechem[1], Robert A. Montgomery[3], Alexandre Loupy[2]

1 Bruker Spatial Biology, Seattle, WA
2 Institut de Transplantation et Régénération d'Organes de Paris (PITOR), Université Paris Cité, Paris, France
3 NYU Langone Transplant Institute, New York, NY
4 Université de Poitiers, Poitiers, France

Kidney xenotransplantation is a promising approach that has the potential to open new treatment avenues for the hundreds of thousands of patients currently on transplant waiting lists. Here we sought, for the first time, to characterize the long-term immune response to the transplantation of a genetically modified porcine kidney into a decedent human recipient. We performed in-depth multimodal phenotyping of tissue biopsies taken at seven sequential follow-up times from Day 0 through Day 61. Initial bulk transcriptomic studies with the nCounter® Human Organ Transplant panel showed a molecular signature evocative of antibody-mediated rejection developing as early as day 10. In addition to a pre-transplant pig kidney control sample, we selected four post-transplantation time points to further explore and used the CosMxTM Spatial Molecular Imager to localize and profile the human immune cells infiltrating the xenograft in their spatial context. Without any custom transcript additions to the Human CosMx 6K Discovery panel, we designed a novel bioinformatic approach to accurately distinguish the approximately 3,000 human immune cells from the approximately 70,000 pig structural cells with high confidence. Across all time points, macrophages were the most prevalent human cell type, in agreement with previous analyses. The human immune rejection response peaked at the D33 time point after which anti-rejection treatment was provided, leading to a successful decrease in the immune-molecular signs of antibody-mediated rejection. We observed human myeloid cells present in every one of the post-transplant glomeruli in the sampled biopsies and characterized differential expression of pro- and anti-inflammatory myeloid genes over the course of treatment. We identified spatially-enriched human macrophage – pig structural ligand-receptor interactions present at early and late timepoints and catalogued spatially correlated modules of co-expressed genes. Overall, our data suggest xenoantibody-mediated rejection, characterized by human innate immune cells infiltration in the porcine xenograft, localized within and outside of glomeruli. Discovery of these molecular mechanisms and the successful attenuation of the xenograft organ rejection with targeted treatment paves the way for potential new anti-rejection therapies to further enable human clinical trials.

Connecting form and function: mapping microglial spatial biology in a mouse model of acute and chronic ischemic stroke

Kimberly Young[1], Ashley Heck[1], Alyssa Rosenbloom[1], Aster Wardhani[1], Maximilian Walter[1], Rachel Liu[1], Lidan Wu[1], Claire Williams[1], John Lyssand[1], Christine Kang[1], K.P. Doyle[2], Helena W. Morrison[3], Margaret Hoang[1], Joseph M. Beechem[1]

1 Bruker Spatial Biology, Seattle, WA
2 University of Arizona College of Medicine, Tucson, AZ
3 University of Arizona College of Nursing, Tucson, AZ

Microglia, the brain's primary immune cells, are highly responsive to changes in their environment, adapting their behavior through intricate molecular signaling. These functional shifts, from surveillance to injury response, are associated with morphological changes, which can serve as markers of their functional state. Our study uniquely combines cutting edge, high-plex spatial proteomics and traditional immunohistochemistry (IHC) to quantify and correlate microglial morphology with functional changes after ischemic stroke at different timepoints.

Adult male mice were subjected to transient middle cerebral artery occlusion and reperfusion using the filament method. Following this procedure, stroke severity was confirmed using a Bruker Biospec 70/20 7.0T MRI scanner. Brains were extracted at 24 hours, 2 weeks, or 4 weeks post-stroke (3 animals/timepoint), fixed, and sectioned at 10µm (spatial proteomics) or 50µm (IHC). We stained microglia with an IBA1 antibody or, utilizing the high plex Mouse CosMx® Neuroscience Protein Panel and the CosMx Spatial Molecular Imager (SMI), we detected 68 proteins with single cell resolution, focusing on regions proximal and distal to the infarct.

Leveraging the many microglial and immune markers in the Mouse CosMx Neuroscience Protein Panel, we captured precise cell morphology and quantified it by skeleton analysis focused on microglial process length and endpoints. Our results showed strong correlation between CosMx SMI and IHC morphometric data over time and distance from stroke injury, supporting CosMx SMI as a viable alternative to traditional IHC. Additionally, principal component analysis of select proteins in the CosMx SMI data revealed correlations between microglial form and function across various brain regions. Utilizing the InSituCor analysis toolkit, we further identified a module of spatially correlated proteins in the infarct region two weeks post-stroke, including Cathepsin B—a promising therapeutic target—and phagocytic markers CD11b, CD11c, and DAP12, with their colocalization shifting gradually with distance from the infarct core.

By pairing spatial proteomic data with microglial morphology analysis, we gained deeper insights into the complex interactions between microglia and the recovering brain following ischemic stroke, highlighting how these interactions evolve over time and with proximity to the infarct core.

Expanding the diagnostic yield of sensorineural hearing loss through whole-genome sequencing: a comprehensive genomic approach

Sangmoon Lee[1], Sang-Yeon Lee[2,3,4*], Seungbok Lee[2,5*], Seongyeol Park[1*], Sung Ho Jung[3], So Min Lee[3], Won Hoon Choi[2], Yejin Yun[2], Ju Hyuen Cha[2], Hongseok Yun[1], Sangmoon Lee[5], Myung-Whan Suh[2], Moo Kyun Park[2], Jae-Jin Song[6], Byung Yoon Choi[6], Jun Ho Lee[3], Young Seok Ju[1,7], Seung Ha Oh[3], June-Young Koh[1], Erin Connolly-Strong[1], Jong-Hee Chae[2,5]

1 Inocras Inc., San Diego, CA
2 Seoul National University Hospital, Department of Genomic Medicine, Seoul, South Korea
3 Seoul National University College of Medicine, Seoul National University Hospital, Department of Otorhinolaryngology, Seoul, South Korea
4 Seoul National University Medical Research Center, Sensory Organ Research Institute, Seoul, South Korea
5 Seoul National University Children's Hospital, Seoul National University College of Medicine, Department of Pediatrics, Seoul, South Korea
6 Seoul National University Bundang Hospital, Seoul National University College of Medicine, Department of Otorhinolaryngology, Seongnam, South Korea
7 Korea Advanced Institute of Science and Technology, Graduate School of Medical Science and Engineering, Daejeon, South Korea
*These authors contributed equally to this work.

Background:
Sensorineural hearing loss (SNHL) is a condition caused by damage to the inner ear or auditory nerve, often with a genetic basis. Despite its prevalence, the genetic underpinnings of SNHL remain incompletely understood. In some cases, SNHL may be a symptom of a rare underlying disease, adding complexity to the diagnosis. As with many rare diseases, identifying the genetic components of SNHL can be challenging, often leading to prolonged diagnostic journeys and delayed treatment. Comprehensive genomic analysis is critical to uncover the causes of SNHL in both common and rare disease contexts.

Methods:
This study included 394 unrelated patients diagnosed with SNHL, along with their family members. Patients were categorized into non-syndromic SNHL (n=342) and syndromic SNHL (n=52). A structured, multi-step genetic testing approach was used. Initially, all non-syndromic patients underwent RT-PCR with classical deafness genes. Undiagnosed patients (n=294) progressed to targeted panel sequencing (n=99) or whole-exome sequencing (WES) (n=249). Further testing included mtDNA sequencing (n=11) and multiplex ligation-dependent probe amplification (MLPA) (n=54). Finally, whole-genome sequencing (WGS, RareVision, INOCRAS, San Diego, CA) was performed on 120 patients who remained undiagnosed after earlier tests.

(continue to page 122)

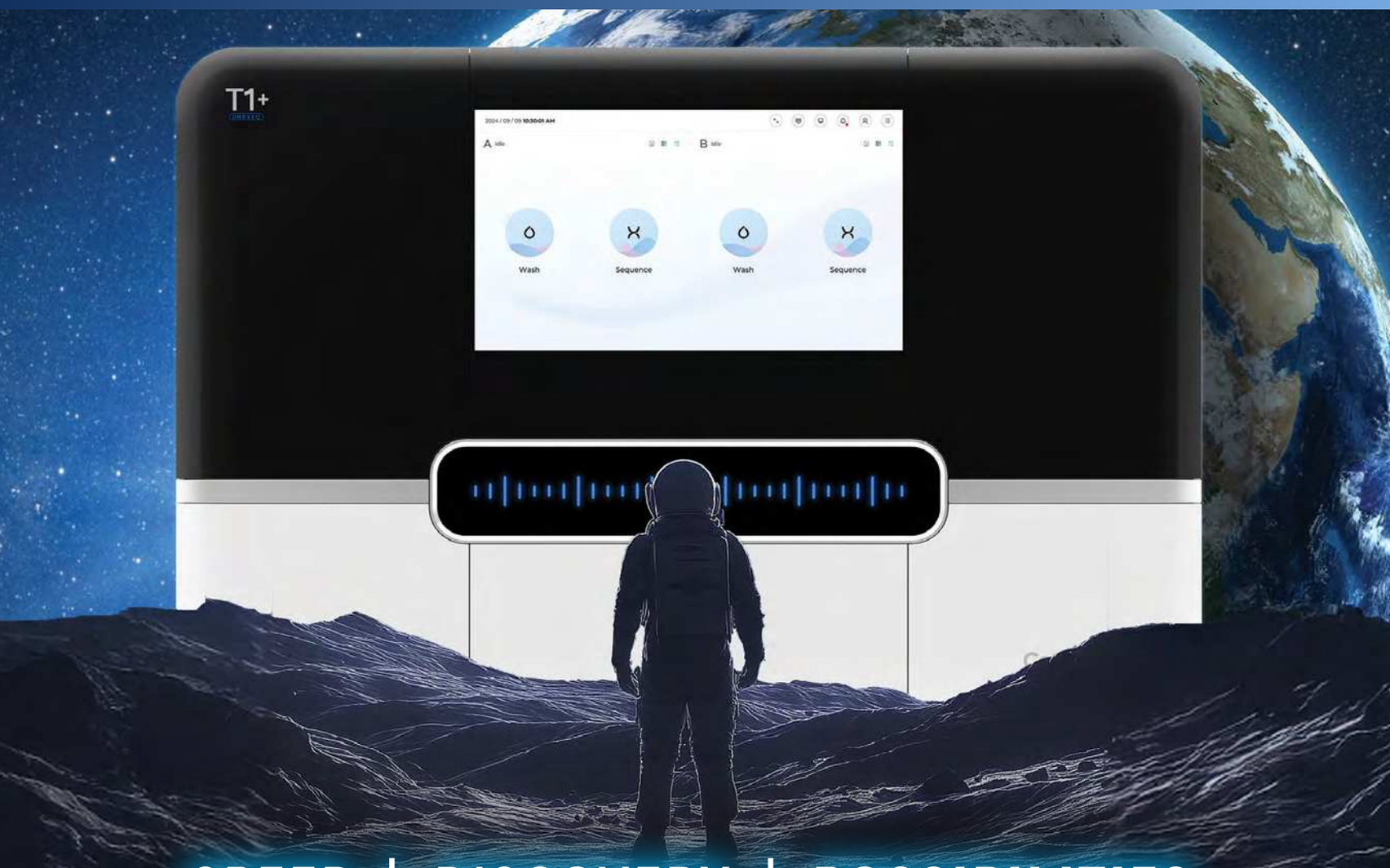
(continued from page 121)

Results:
The overall diagnostic yield was 55.8%, with 220 of 394 patients receiving a genetic diagnosis. Of the 120 patients who underwent WGS, 24 were diagnosed, representing a diagnostic yield of 20% in this group. Many of the causal variants identified through WGS were variants less likely to be detected by previous methods, such as single nucleotide variants (SNVs) in deep intronic regions and structural variants (SVs).

Conclusion:
These findings demonstrate the value of WGS in identifying rare genetic variants that may be missed by targeted sequencing approaches. The inclusion of WGS significantly enhances diagnostic yield, especially for complex cases of SNHL, facilitating more timely and accurate interventions.

DISCOVER YOUR NEW WORLD OF GENOMIC RESEARCH

DNBSEQ-T1+ SEQUENCER



SPEED | **DISCOVERY** | **POSSIBILITIES**

JOIN THE EXPLORATION WITH DNBSEQ-T1+ AT OSPREY 6!

BRONZE 4 SPONSOR WORKSHOP | TUESDAY 3:20 – 3:35 PM

Learn more about AGBT events
completegenomics.com/AGBT2025

Learn more about DNBSEQ-T1
completegenomics.com/DNBSEQ-T1-plus

Complete GENOMICS™

Technology and Application Development

Bridging the gap in high-molecular-weight DNA extraction: scalable automation using digital fluidics

Nayan Agarwal[1], Daniel Giguere[1], Maham Zia[1], Shikhar Gupta[1], Abdul Majeed Mohammed[1]

1 Volta Labs, Boston, MA

Long-read sequencing technologies, such as PacBio and Oxford Nanopore Technologies, are transforming genomic research by enabling complete, telomere-to-telomere assemblies. A critical prerequisite is high-molecular weight (HMW) DNA, which is necessary to capture complete repetitive genomic regions in single, contiguous reads. Traditionally, generating HMW DNA has required either extensive training or expensive automated platforms designed for high-throughput, single-purpose workflows. These high-capacity systems (typically 96 samples per batch) are inaccessible and often not required for labs with moderate sample volumes, creating a bottleneck in the adoption of these advanced sequencing techniques as the cost of sequencing continues to drop.

To bridge the gap between throughput and automation, Volta Labs has developed a comprehensive solution for HMW DNA extraction. The VoltaPure™ HMW DNA Extraction App provides all reagents and consumables needed and is paired with pre-optimized workflows to operate on the Callisto™ Sample Prep System, eliminating the method development that is traditionally required on automation platforms. Callisto also accommodates flexible batch sizes up to 24 samples. This enables users to load the instrument in less than 30 minutes and walk away, returning to collect high-quality HMW DNA for a variety of sample types and throughputs.

In this study, we evaluated Callisto's performance against manual extraction using several kits (PacBio NanoBind, NEB Monarch HMW extraction) using human whole blood, cell lines, and buffy coat as input. We assessed DNA integrity using Femto Pulse and constructed PacBio HiFi libraries, which were then sequenced on the PacBio Revio. The results demonstrate that Callisto-extracted HMW DNA retains very long DNA fragments without shearing, performing equivalently to manually extracted samples for yield, fragment size, and sequencing metrics. Sequencing metrics evaluated include read quality, read length, and variant calling metrics.

Callisto simplifies HMW DNA extraction into a push-button, walk-away workflow, without compromising on read lengths or quality. HMW DNA extracted on Callisto is suitable for many downstream applications that require long reads, including genome assembly and variant calling. Callisto enables end-to-end automation of HMW DNA extraction for NGS workflows with a single benchtop instrument, providing a full sample-to-answer solution.

A hitchhiker's guide to long-read RNA isoform sequencing

Aziz Al'Khafaji[1], Brian Haas[1], Akanksha Khorgade[1], Christophe Georgescu[1], Houlin Yu[1], Ghamdan Al-Eryani[1], Dan Bartlett[1], Emily White[1], Asa Shin[1], James Webber[1], Victoria Popic[1]

1 Broad Institute of MIT and Harvard, Broad Clinical Labs, Cambridge, MA

Recent advances in long-read RNA isoform sequencing have enabled researchers to resolve the rich complexity of alternative splicing in the transcriptome at scale. This new capability naturally comes with a set of challenges; new tools must be developed for data processing, QC, analysis, and visualization. Further, appropriate computational infrastructure must be built to reproducibly benchmark latest methods and derive principled best practices. Finally, robust and easy to use pipelines are needed to enable non-expert users to leverage these powerful methods. To this end, we have developed a suite of tools, best practices, and vignettes hosted on a publicly available website to guide researchers on their journey from RNA material to fully analyzed data.

Through guided examples, we present the use of our latest suite of tools and existing methods to make the most of long-read RNA-seq data. RNA isoform library construction: Our resource begins with established methods for cDNA synthesis, best practices and common pitfalls which result in artifacts and incomplete capture of mRNAs. Sequencing platforms: Next, we review approaches for scaled cDNA sequencing, identifying sequencing platform tradeoffs. Quality control: After data is generated, we implement read level QC tool sets, MARTI and LR-RNAQC+. MARTI is a read classification tool that identifies adapter composition and classifies reads as coming from a properly formed cDNA or as artifactual with discrete adapter compositions. After classification, LR-RNAQC+computes and plots key quality metrics, including read-level SQUANTI classifications. Single-cell isoform sequencing: Leveraging our QC tools, we characterize 10x and Fluent single-cell platforms for isoform sequencing and highlight open challenges in the full-length mRNA capture and transcript quantification. Isoform ID benchmarking: Next, we review our benchmarking infrastructure and highlight leading isoform identification and quantification tools in the field. Analysis: Transitioning to analysis, we present a vignette showing users how to perform differential isoform expression analyses and visualizations using Integrated Transcriptome Viewer (ITV). Finally, we conclude the presentation demonstrating the power of this data type; simultaneously resolving transcript isoforms, fusions, and mutations in single-cell data from >200,000 cells, 54 patients over the course of a breast cancer clinical trial.

Multi-omic genomic mapping with long-read sequencing

Bryan J. Venters[1], James Anderson[1], Paul W. Hook[2], Jamie Moore[2], Morgan Oatley[1], Vishnu U. Sunitha Kumary[1], Allison Hickman[1], Anup Vaidya[1], Sabrina Hunt[1], Ryan Ezell[1], Jonathan M. Burg[1], Zu-Wen Sun[1], Martis W. Cowles[1], Winston Timp[2], Michael-Christopher Keogh[1]

1 EpiCypher Inc., Durham, NC
2 Johns Hopkins University, Department of Biomedical Engineering, Baltimore, MD

Gene transcription is regulated by the complex interplay between histone post-translational modifications (PTMs), chromatin associated proteins (CAPs), and DNA methylation (DNAme). Mapping their genomic locations and examining the relationships between these chromatin elements is a powerful approach to decipher mechanisms of disease, thereby enabling discovery of novel biomarkers and therapeutics. Leading epigenomic mapping technologies (e.g., ChIP-seq, CUT&RUN) rely upon DNA fragmentation to isolate regions of interest for sequencing on short read platforms (e.g., Illumina). This strategy leads to substantial loss of contextual information regarding the surrounding DNA, precluding the identification of multiple co-occurring epigenomic features on a single DNA molecule. By contrast, long-read sequencing (LRS) platforms are capable of sequencing very long reads from a single molecule (typically >10kb), allowing relationships between features on a single molecule to be used to resolve heterogeneity within mixed populations.

Here we report a robust multi-omic method that leverages LRS to simultaneously profile histone PTMs (or CAPs), DNAme, and parental haplotype in a single assay. This nondestructive, epigenomic mapping approach leverages a novel DNA methyltransferase fusion protein (pAG-M.EcoGII) to label DNA underneath antibody-targeted chromatin features, thereby marking sites of interest while preserving DNA molecules intact for LRS. Inspired by our work with state-of-the-art immunotethering-based approaches (CUT&RUN / CUT&Tag), nuclei are bound to magnetic beads to streamline and automate sample processing. Next, adenosines nearby antibody-targeted chromatin features are methylated with pAG-M.EcoGII, which are then directly read from genomic DNA using Oxford Nanopore Technologies or Pacific Biosciences LRS platforms. To determine the capabilities and limitations of this assay, we tested multiple chromatin targets in various cell lines. Importantly, this method is highly reproducible across biological replicates, and highly concordant with orthogonal SRS assays (e.g., CUT&RUN). Further, we showed that this method is a true multi-omic approach by simultaneously profiling histone PTMs, native DNAme (5mC), and parental single-nucleotide variants from single DNA molecules within a single reaction. Finally, this workflow preserves chromatin integrity for LRS, revealing heterogeneity (e.g., haplotype or paternal origin) within / between data types and providing access to previously unmappable genomic regions (e.g., centromeres).

Targeted genomic and epigenomic dual analysis for the same region of a single gene

Yun Bao[1], Grace Zhao[1], Heng Wang[1], Ashraf Wahba[1], Yingmin Wang[1], Bilan Hsue[1], Ling Su[2], Josh Wang[1], Shengrong Lin[1], Xiaolin Wu[2]

1 Agilent Technologies, Santa Clara, CA
2 Frederick National Laboratory for Cancer Research, Frederick, MD

Omics analysis for sensitive cancer detection and treatment stratification is increasingly focusing on integrating genomic and epigenomic data. While most studies often limit methylation analysis to promoter regions and genomic analysis for exons, research suggests the importance of examining omics information for entire genes, including UTRs and introns. Current technologies at the whole genome level for simultaneous genomic and epigenomic analysis require costly deep sequencing for low tumor content and may suffer from limited coverage in high CG content regions.

To address these challenges, we developed a novel strategy that enables genomic alteration and methylation dual analysis of a same region within a single gene from both gDNA and cfDNA. This approach offers an affordable solution with improved uniformity including the high CG regions.

Using ARID1A as a demo case, we developed a novel strategy for simultaneous genomic and epigenomic analysis using the Agilent Avida Duo target enrichment system. We were able to sequentially capture and obtain both genomic and methylation information.

To optimize sequencing budget and coverage uniformity, we divided the panels (90kb) into high-quality and low-complexity regions and experimented with different panel mixture ratios. With 10ng fragmented gDNA, we achieved good coverage for both genomic (>1000x unique molecules) and methylation (>100x unique molecules) analysis, using less than 5M paired-end sequencing reads for each in the Avida Duo workflow. For uniformity, the percentage greater than 0.2x of mean coverage is > 90%.

We also compared two fragmentation methods (fragmentase and Covaris sonication) and obtained comparable results for both genomic alteration and methylation analysis.

Our pilot runs on tumor gDNA, FFPE, and cfDNA samples (n>=5) successfully demonstrated the feasibility of our approach for simultaneous targeted genomic and epigenomic analysis of the same region in a single gene with low input DNA requirement.

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

A comparison of 10x Genomics 5' HT kit versus 5' GEM-X kit with 'super-loading' to enable cost-effective single-cell sequencing of millions of PBMCs using a new automated approach

Sarah Cooper[1], Abby Crackett[1], Lesley Shirley[1], Cristina Cotobal Martin[1], May Hu[1], Marcus Tutert[1], Matiss Ozols[1], Vivek Iyer[1], Aleksander Gontarczyk[1], Stephen Watt[1], Ian Johnston[1], Carl Anderson[1]

1 Wellcome Sanger Institute, Hinxton, Cambridge, United Kingdom

The field of human genetics has identified tens of thousands of genome regions significantly associated with human complex diseases and traits. To translate these findings to actionable therapies the next major challenge is to identify the genes, cell types and pathways perturbed by these associated variants. To address this, Wellcome Sanger Institute (WSI) is developing a robust automated workflow to generate single-cell RNA sequencing data across tens of thousands of Peripheral Blood Mononuclear Cells (PBMC) samples, including upfront PBMC isolation and banking from whole blood.

Here we present data from the evaluation of two 10x Genomics kits (5' HT kit and 5' GEM-X kit) within our automated pipeline. In both instances we have been 'super-loading' the chips beyond the standard 10x Genomics protocols in order to enable cost-effective single-cell sequencing of PBMCs. Using our miniaturised automated 96-well workflow we isolated, counted and banked PBMCs from 10 donors. After thawing, counting and pooling the cells with new automated workflows, we manually loaded the PBMCs onto the 10x Genomics 5' HT and 5' GEM-X chips in parallel to carry out our benchmarking experiment. We found that the GEM-X kit had higher cell recovery than the HT kit (56% versus 45.9%) and we were able to push the number of cells recovered on both kits to around 50K per lane. Both kits recovered the same proportion of donors and proportion of different cell types. Importantly we found that the GEM-X kit produced higher quality data, and we were able to detect a higher number of genes per cell.

This work has enabled us to determine the most suitable technology for future use, GEM-X, as well as maintaining a workflow that is flexible to future changes and the ability to process other library types as necessary (e.g. long-read sequencing of full-length cDNA, CITE-seq and TCR/BCR).

Benchmarking of biomodal evoC 6-base sequencing versus Illumina Infinium DNAm arrays

Kenneth Beckman[1], Mark Murphy[1], Nick Jones[1], John Garbe, Noah Strom[1]

1 University of Minnesota Genomics Center (UMGC), Minneapolis, MN

The genome-wide analysis of DNA methylation at CpG (5mC) sites can be achieved by several methods, with array-based approaches from Illumina having dominated the analytical space for two decades, and sequencing-based approaches (WGBS, RRBS) becoming common as NGS costs have dropped. The nature of the information provided by DNAm arrays is strikingly different from NGS. Compared to arrays, which interrogate tens of thousands of molecules and provide CpG site beta values (5mC/C) with high precision, NGS typically interrogates tens of reads at a given locus, with single-site 5mC/C precision limited by read depth. At the same time, compared to NGS, which provides allelic information on neighboring CpG sites found within a single read and reports on all mappable CpG sites in the genome, arrays mask allelic information and target a small fraction of CpG sites. Until recently, moreover, neither arrays nor NGS have allowed for the facile genome-wide measurement of hydroxymethylcytosine (5hmc) along with 5mC. With the introduction of the duet multiomics solution evoC from biomodal, genome-wide 6-base sequencing (A, C, T, G, 5mC, 5hmC) is possible in a single workflow. Consequently, we have benchmarked biomodal's evoC to Illumina's Infinium HumanMethylationEPIC v2 and MethylationScreeningArray-48 in order to provide bi-directional insights into the platforms. We find that in terms of CpG site beta values, the platforms are well correlated (with quantitative agreement limited, as expected, by sequencing depth), bolstering confidence in both. Our comparison reveals that arrays are less able to distinguish extremes of methylation state (100% methylated or 0% methylated) than evoC and suggests that by leveraging allelic information from neighboring CpG sites, the precision of lower depth sequencing to deliver a comparable beta value to arrays may be improved. We have modeled the use of evoC to complement arrays in terms of performance, cost, and utility in epigenetic clock analysis, and have attempted to generate a list of array CpG site reliability using evoC data as a truth model.

Advancements in in vitro sequencing: automated, high-throughput insights into cellular dynamics and gene function

Pedro Belda Ferre[1], David White[1], Jake LeVieux[1], Connor Thompson[1], Kousik Sundararajan[1], Mark Ambrosio[1], Jacob Moreno[1], Chris Benitez[1], Michael Klein[1], Ram Adhikary[1], Andy Jo[1], Sinan Arslan[1], Matthew Kellinger[1], Michael Previte[1]

1 Element Biosciences, San Diego, CA

The emergence of in vitro sequencing technologies marks a significant advancement in our capacity to investigate biological systems with unparalleled resolution. Traditional sequencing approaches require the extraction and amplification of nucleic acids, which can introduce biases and obscure spatial context. In contrast, in vitro sequencing facilitates the direct analysis of nucleic acids within their native environments, yielding valuable insights into cellular dynamics, host-pathogen interactions, and the functional roles of specific genes in vivo. However, current in vitro sequencing methods face challenges, including manual cycling workflows and a high background noise that accumulates over the course of the experiment. These limitations restrict read length, quality and throughput. Despite these hurdles, in vitro sequencing offers substantial promise for applications such as optical pooled screening (OPS) and T-cell receptor (TCR) sequencing, allowing for the visualization and analysis of complex biological interactions. In our study utilizing the AVITI24 platform, we demonstrate that transcripts can be directly sequenced using both targeted and untargeted approaches. Up to 12 cell populations can be cultured directly on the sequencing flowcell, followed by automated 1x100 base pair sequencing using avidity chemistry. By targeting the 3'UTR of cellular transcripts in HeLa cells, we detected expression of 10,868 human genes with 98% of aligned reads on-target. Using a targeted approach, we were able to directly sequence transcripts of interest to reveal heterogeneity of transcript content within a single cell and across a cellular population. These targeted and untargeted methodologies enable automated workflows for in vitro sequencing applications, achieving 100 cycles of sequencing on up to 600,000 cells. Furthermore, we demonstrate an expansion of this method to triple sequencing output by implementing 3x100 sequencing, effectively enhancing the information density per cell. These innovations provide the first widely available, customizable solution for automating sequencing within cells.

An integrated short- and long-read multi-omics approach to resolve the cellular diversity of ALS human spinal cord at single-cell resolution

Shirin Issa Bhaloo[1], Rui Fu[1], Thomas Chu[1], Tanya Hemrajani[1], Maria Hauge Pedersen[1], Archana Yadav[2], Jason A. Mares[2], Vilas Menon[2], Hemali Phatnani[1,2], Jade Carter[1]

1 New York Genome Center, New York, NY

2 Columbia University Irving Medical Center, New York, NY

The central nervous system (CNS) is composed of a complex network of cellular- heterogeneity. Spinal motor neurons (MNs) specifically degenerate in patients with amyotrophic lateral sclerosis (ALS), while activated microglia have been implicated in ALS pathogenesis(1, 2, 3). To elucidate the molecular basis for this selective phenotype, it is crucial to characterize the repertoire of cells in the CNS and their respective roles in health and disease. Single-cell technologies have emerged as powerful tools to enhance our understanding of biology and disease. Single-cell RNA-sequencing and spatial transcriptomics studies have been conducted to characterize the underlying transcriptional molecular mechanisms in ALS(1, 4). However, a multi-omics-based approach at single-cell resolution would provide deeper insights into comprehensively deciphering the molecular pathogenesis in the CNS.

We developed an integrated short and long read single-cell multi-omics approach on frozen postmortem human spinal cord (SC) tissue from ALS patients. First, a nuclei isolation method was optimized for frozen human SC. To simultaneously interrogate cell-type-specific gene expression and chromatin accessibility, the 10X Genomics Multiome protocol was applied and sequenced on the NovaSeq X. Cell clustering and cell-type mapping showed that cell profiles, independently derived from both gene expression (RNA) and gene activation (ATAC), identified all the 11 major known human SC cell classes, including the rare population of MNs.

To identify cell-type specific isoforms from the same nuclei captured in short read analysis, long read RNA libraries were prepared using the cDNA from the Multiome protocol and sequenced on a PromethION. Resulting cell numbers and cell-type mapping matched that which was observed from short read Multiome analysis, albeit with an expected lower output of reads/cell on the long read sequencing platform. Further analysis showed cell-type specific isoform switching, thus shedding light on the dynamics of isoform expression variation in ALS pathogenesis.

Here, we present a comprehensive multi-omics approach at single-cell resolution on human postmortem spinal cord. The integration of gene expression coupled to chromatin accessibility signatures with isoform switching profiling allows for a more robust capture of the functional state and identity of individual cells within a complex tissue.

Ultra-throughput targeted sequencing using pre-indexed molecular inversion probes

Tamir Biezuner[1], Adi Danin[1], Nelly Frenkel[1], Adi Ben-Yehuda[1], Reut Nuri[1], Michael Gershovits[1], Yardena Brilon[1], Hadar Haham[1], Noa Tzunz-Henig[1], Brian Ross[1], Tom Fleischer[1]

1 Sequentify Ltd., Rehovot, Israel

Despite advances in sequencing technologies, Next-Generation Sequencing (NGS) library preparation now faces new hurdles as it scales up for broader adoption in healthcare. The labor-intensive, time-consuming and costly nature of sample preparation has become a significant bottleneck, highlighting the need for efficient solutions capable of supporting scalable and decentralized workflows. To address these challenges, we developed InfiniSeq, a pre-indexed Molecular Inversion Probe (MIP) technology. It builds upon the familiar four-step MIP library preparation protocol but utilizes pre-indexed probe panels. Upon probe hybridization, each reaction is barcoded and can immediately be pooled with other pre-indexed reactions, allowing processing as a quick (3.15-hour) single reaction from the second step onwards. This facilitates high-throughput experiments in a single-pot reaction.

We validated the method by running 24 clinical samples individually using the standard MIP reaction and collectively using the InfiniSeq reaction in a single pooled manner. The results demonstrated the ability to detect mutations at low variant allele frequencies (VAF) down to 1%, with no discernible cross-contamination. These findings not only aligned with anticipated mutation signatures but also exhibited consistent performance in key metrics such as on-target rate and uniformity. Further experiments, using synthetic and biological samples, demonstrated the ability to pool up to 96 pre-indexed reactions for genotyping purposes without an increase in cross-reactivity between samples.

InfiniSeq aims to facilitate NGS's next phase in healthcare by reducing barriers to entry and enabling labs with minimal infrastructure to implement scalable NGS workflows efficiently, paving the way for broader integration and impact of NGS in clinical settings.

Streamlining imaging-based spatial biology on CellScape with EpicIF, a universal fluorophore removal strategy

Oliver Braubach[1], Jannik Boog[1], Thore Böttke[1], Arne Christians[1]

1 Bruker Spatial Biology, Saint Louis, MO

Imaging-based spatial biology is an iterative process of biomarker labeling, biomarker imaging, and biomarker signal removal. The process is repeated until the desired plexity of imaging data is achieved. The removal of fluorescence signals after each detection cycle is central to the experiment but is pursued in various ways, including photobleaching of fluorophores, chemical stripping of antibodies, as well as the use of barcoded antibodies. Each signal removal method has limitations and introduces unique challenges to the realization of multiplex assays. To resolve assay-based limitations, and to streamline the integration of multiplex experiments, we developed Enhanced Photobleaching in Cyclic Immunofluorescence (EpicIF). EpicIF allows quick, gentle, and complete removal of common photostable fluorophores in situ, and enables rapid development and customization of novel spatial omics assays on the CellScapeTM spatial biology platform.

We first demonstrated the ability of EpicIF to remove signals from common fluorophores in vitro. Our data show that EpicIF rapidly decolorizes essentially all commercially available photostable organic dyes, including rhodamine, cyanine and BODIPY-based dyes. Tested in situ, EpicIF maintained speed and universality of signal removal in multiplex immunofluorescence (mIF) experiments, regardless of biomarker plex levels and tissue types tested. Since EpicIF requires no chemical modifications of target antibodies or experimental tissues, it presented an unprecedented opportunity to integrate multi-omic analyte detection in same tissue sections. To this end, we successfully integrated mIF with the detection of RNA targets via hybridization chain reaction RNA-FISH (HCR RNA-FISH), as well as the detection of protein-protein interactions via in situ proximity ligation (isPLA), essentially a tri-omic same section experiment.

Our data show that EpicIF on CellScape allows for fast, virtually unlimited and modular multiplexing of imaging-based spatial biology experiments. Our method allows signal removal of photostable fluorophores, independent of which biomarker assay is deployed, and thereby opens the door for the development and rapid deployment of a wide range of true multi-omic multiplex imaging methods.

Beyond the genome: advancing our understanding of the proteome with Next-Generation Protein Sequencing™

Meredith L. Carpenter[1], Mathivanan Chinnaraj[1], Kendrick Nguyen[1], Kenneth Skinner[1], Ben Moree[1], Brian Reed[1], John Vieceli[1]

1 Quantum-Si, Branford, CT

Protein sequencing is an emerging proteomics tool that enhances genomics and transcriptomics research by revealing key features of proteins encoded by the genome, thereby bridging the gap between genetic information and biological function. In this study, we explored how the benchtop Quantum-Si Platinum® instrument, the first Next-Generation Protein Sequencer™, offers an integrated understanding of cellular processes by detecting changes at the protein level (such as post-translational modifications, or PTMs) that cannot be captured by genomics data alone.

In the Next-Generation Protein Sequencing (NGPS) workflow, individual peptides are immobilized on a semiconductor chip, probed by dye-labeled N-terminal amino acid (NAA) recognizers, and then subjected to aminopeptidase cleavage to expose each subsequent NAA for recognition. The fluorescence features and kinetics of each NAA binding event are recorded and converted into amino acid calls. This information-rich datatype with single amino acid resolution enables detection of multiple types of proteomic variation that can be challenging to recover using standard proteomics approaches.

To explore the versatility of this technology, we first evaluated the detection of single amino acid variants. We observed that NGPS discriminated three variants of SARS-CoV-2 spike protein and two variants of ubiquitin with single amino acid substitutions. Next, we tested the ability of NGPS to identify PTMs by measuring citrullination and dimethylation of arginine—two PTMs that play key roles in disease. NGPS was able to discriminate these PTMs via differences in recognizer binding. To evaluate identification of unknown proteins, we isolated proteins directly from human serum and correctly identified them using an 8,000-protein reference panel. Finally, we developed a novel peptide barcoding method to screen and characterize proteins, with applications in mRNA vaccine development, protein engineering, and antibody engineering.

Taken together, our data demonstrate the capacity of NGPS to detect a range of amino acid variants, identify unknown proteins, and enable high-throughput screening. Therefore, Platinum has the potential to become a key component in multi-omics and drug development workflows.

Unlocking multiomic analysis in spatial genomics: integrating spatial ATAC-seq and V(D)J sequencing with gene expression using novel spatial barcoding of single nuclei

Technology and Application Development

POSTER 413

Christina Chang[1], Julie Wilhelmy[1], Wanxin Wang[1], Christina Fan[1]

1 Curio Bioscience, Palo Alto, CA

While single-cell multiomic analysis such as RNA-seq, ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing) or V(D)J sequencing have advanced our understanding of gene expression, chromatin accessibility, and immune receptor diversity, these methods miss the critical spatial organization of cells within their native environments. Existing spatial methods, whether it's microscopy or sequencing-based, primarily focus on gene or protein expression and often lack true single-cell resolution, relying on complex cell segmentation or deconvolution methods. This creates a significant challenge in studying the regulatory landscape and immune receptor diversity in the spatial context of tissues.

To address this gap, Curio Trekker, based on the Slide-tags technology, offers a novel solution by transforming any single-cell sequencing assay into spatially resolved single-cell data. This technology utilizes unique DNA barcodes on a tightly packed bead array to spatially label individual cells. The resulting tagged nuclei are then processed on conventional single-cell sequencing platforms, preserving spatial information while providing true single-cell resolution.

In this study, we apply Curio Trekker to achieve spatial ATAC-seq combined with gene expression analysis, as well as spatial V(D)J sequencing paired with gene expression. These previously difficult-to-achieve combinations enable the exploration of chromatin accessibility and gene regulation within the tissue's native spatial context, as well as spatial mapping of immune receptor diversity and functionality. This integrated spatial multiomics approach provides a more comprehensive view of tissue microenvironments.

We validated these approaches across immune and tumor tissues, demonstrating the robustness and versatility of Curio Trekker in capturing high-resolution spatial data. Our findings highlight how this technology can unlock new insights into cellular regulatory mechanisms and immune cell interactions, advancing research in immunology, oncology, and other fields.

Infinium™ Methylation Screening Array: novel array for epigenome-wide association studies

Evgenii Chekalin[1], David C. Goldberg[1], Cameron Cloud[1], Sol Moe Lee[1], Bret Barnes[2], Steven Gruber[2], Elliot Kim[1], Anita Pottekat[2], Maximillian S. Westphal[2], Luana McAuliffe[2], Elisa Majounie[2], Manesh Kalayil Manian[2], Qingdi Zhu[2], Christine Tran[2], Mark Hansen[2], Jared B. Parker[3], Rahul M. Kohli[3], Rishi Porecha[2], Nicole Renke[2], Wanding Zhou[1],[4,*]

1 The Children's Hospital of Philadelphia, Center for Computational and Genomic Medicine, Philadelphia, PA
2 Illumina, Inc., San Diego, CA
3 University of Pennsylvania, Department of Medicine, Philadelphia, PA

It is well known that DNA methylation is affected by environmental factors. For over a decade, the quantification of DNA methylation in humans has been enabled by Epigenome-Wide Association Studies (EWAS), based on Illumina Infinium Methylation450K and EPIC BeadChips. In recent years, there has been increasing interest to scale EWAS to the population level. However, large cohort studies can be costly with existing tools. Here we present a novel array designed to democratize and scale epigenetics research.

The Infinium Methylation Screening Array includes more than 272K unique methylation sites and is roughly twice as affordable as the InfiniumEPICv2.0 BeadChip per sample. Approximately half the content targets known biologically relevant sites for EWAS, based on over a decade of trait associations discovered from the field. The remaining 50% of content covered is obtained from public bulk and single-cell WGBS datasets, targeting cell type atlases, gene expression, mono-allelic expression, and CpGs shown to be 5-hydroxymethylated at the pan-tissue and tissue-specific level.

We comprehensively evaluated this array's reproducibility, accuracy, and capacity for cell-type deconvolution and supporting 5-hydroxymethylation profiling in diverse human tissues. Altogether, the newly developed Methylation Screening Array is a promising tool that enables population-level methylation studies.

Massively scalable direct guide RNA capture for pooled CRISPR screening

Technology and Application Development

POSTER 415

Kristina Fontanez[1], Hannah Rickner[1], Robert Meltzer[1], Yi Xue[1], Jacob Ishibashi[1], Aaron May-Zhang[1], Sophia Bevans[1], Christopher D'Amato[1], Pabodha Hettige[1], Trinity Smithers[1], Racheal Komuhendo[1], Lucas Yoder[1], Muntasir Rahman[1], Fiona Kaper[1], Cande Rogert[1], Carolyn Conant[1]

1 Illumina Inc., San Diego, CA

Single-cell direct capture CRISPR screens have the potential to uncover critical biological insight across a broad range of research fields such as oncology, immunology, and more. However, their utility to date has been limited by the availability of a truly high-throughput, economical, and accessible single-cell direct CRISPR capture technology. The number of cells processed per experiment scales with the number of perturbations; this can result in experiments requiring the processing of tens to hundreds of thousands of cells to ensure sufficient representation of each perturbation.

In response to this challenge, we present a novel method enabling simultaneous capture of single-cell transcriptomes and CRISPR-Cas9 guide RNAs from individual cells facilitated by dual-oligo functionalized polymer particles called particle-templated instant partitions (PIPs). With the potential for multimodal profiling of up to one million cells in a single reaction, this method makes possible the profiling of pooled screens with tens of thousands of guides, a significant advance in direct capture CRISPR screening. To demonstrate the utility and scalability of this technology, millions of CRISPR-Cas9 perturbed mammalian cells were combined with dual-oligo PIPs, emulsified by vortexing, and processed to distinct gene expression and guide libraries. We demonstrate the unique ability to archive cDNA on PIPs to enable segregated amplification of gene expression and sgRNA libraries, enhancing specificity and streamlining the workflow. We observed robust direct guide capture among a diversity of samples and confirmed that whole transcriptome mRNA sensitivity was comparable between uni-linker (mRNA only) and dual linker PIPs (mRNA and Cas9). The combination of simple, high throughput single-cell analysis with emerging low cost high-capacity sequencing capabilities will enable routine genome-wide perturbational assays across many biological fields.

LongPlex XL: an improved long-read tagmentation workflow for cost-effective highly multiplexed ⩾12 kb HiFi sequencing libraries

Technology and Application Development

POSTER 416

Maura Costello[1], Christianto Putra[1], Yanyan Liu[1], Zac Zwirko[1], Gavin Rush[1], Joe Mellor[1]

1 seqWell Inc, Beverly, MA

PacBio's Revio™ system enables researchers to generate high-quality long read data at an unprecedented scale. Maximizing the Revio's gigabase output for cost-sensitive applications requires long read lengths of at least 12 kb, but existing methods for preparing SMRTbell libraries of this length have not yet scaled to match the Revio's throughput, remaining both labor- and cost-intensive.

Here, we present an efficient method for generating multiplexed HiFi libraries with read lengths exceeding 12 kb utilizing LongPlexTM, a transposase-based tagmentation method, to simultaneously fragment and barcode genomic DNA in a single step. Enzymatic fragmentation eliminates the need for separate shearing instrumentation, and by barcoding samples at the outset, it enables sample pooling before labor-intensive downstream steps such as size selection and SMRTbell library preparation. To further streamline the workflow, we introduce a novel application of PacBio's Short Read Eliminator (SRE) kit for size selection. While SRE was designed to remove smaller fragments from genomic DNA, we demonstrate its effectiveness in size-selecting DNA after fragmentation, adapter barcoding, and pooling. Performing pooled size selection prior to SMRTbell library preparation provides a significant workflow advantage, allowing for the processing of more samples with reduced time, effort, and reagent costs.

To validate this method, we used LongPlex indexed transposomes to enzymatically fragment and barcode eight replicates of human genomic DNA (500 ng per replicate), achieving a mode size of approximately 20 kb. The eight tagged samples were pooled into a single tube before undergoing SRE size selection, followed by SMRTbell library preparation. Sequencing the pool on the Revio system generated 88 Gb of data, with an average HiFi read length of 14.4 kb. We have since scaled this method to accommodate higher plex pooling, further increasing throughput while reducing costs.

The LongPlex XL approach combines enzymatic fragmentation with pooled size-selection and SMRTbell prep offers a scalable, cost-effective solution for long read library preparation, empowering researchers to fully utilize the Revio's capacity for multiplexed long read applications.

High-throughput assembly of long, complex DNA by mating bacteria

David Craford[1], Takeshi Matsui[1,2], Po-Hsiang Hung[1,3], Han Mei[1,2,4], Xianan Liu[2,5], Fangfei Li[1,3], John Collins[1], Weiyi Li[2,3], Darach Miller[1,2], Neil Wilson[1], Esteban Toro[5], Geoffrey J. Taghon[6], Gavin Sherlock[3], Sasha Levy[1,2]

1 BacStitch DNA, Inc., Los Altos, CA
2 Stanford Linear Accelerator Center (SLAC) National Accelerator Laboratory, Menlo Park, CA
3 Stanford University, Department of Genetics, Palo Alto, CA
4 Asimov, Inc., Boston, MA
5 Twist Bioscience, Inc., South San Francisco, CA
6 National Institute of Standards and Technology (NIST), Gaithersburg, MD

To scale engineering of long complex DNA, we report here a novel in vivo DNA assembly technology platform that combines bacterial conjugation, in vivo DNA cutting, and in vivo homologous recombination to seamlessly stitch blocks of DNA together by mating E. coli in large arrays or pools. For long DNA, this workflow is simpler, cheaper, and higher throughput than current in vitro based DNA assembly approaches such as Gibson Assembly and Golden Gate Assembly that require DNA to be moved in and out of cells at different procedural steps. We perform over 5,000 assemblies with two to 13 DNA blocks that range from 240 bp to 8 kb to generate constructs as long as 23 kb at relatively high throughput and fidelity. Most assembly errors are deletions between long interspersed repeats. However, these errors can be overcome by performing multiple replicates: additional cost and hands-on time for assembly and sequence verification for additional replicates are nominal. We show the platform can be used to build directed combinatorial libraries in arrays or pools, and that DNA blocks onboarded into the platform can be repurposed and reused with any other DNA block in high throughput without a PCR step. Because of these features, the platform has the potential to accelerate design-build-test-learn cycles of DNA products.

CosMx® SMI spatially resolved whole transcriptome in FFPE tissue: a paradigm shift for tissue analysis

Yi Cui[1], Shanshan He[1], Michael Patrick[1], Megan Vandenberg[1], Evelyn Metzger[1], Stefan Rogers[1], Kathy Ton[1], Dan McGuire[1], Haiyan Zhai[1], Margaret Hoang[1], Michael Rhodes[1], Joseph Beechem[1]

1 Bruker Spatial Biology, Seattle, WA

Single-cell whole transcriptome analysis has emerged as a cornerstone of modern biomedical research, offering unprecedented insights into cellular heterogeneity and molecular pathways. Historically, single-cell RNA-sequencing (scRNA-seq) has been the mainstay tool to address this need by enabling the identification of cell types, uncovering cellular states, and mapping signaling networks. Spatial technologies allow the dissection of cells in their native context but lacked the ability to interrogate the whole transcriptome. The CosMx® Whole Transcriptome (WTX) panel, by combining whole transcriptome profiling with nanometer spatial mapping, enables direct analysis and visualization in challenging samples without the limitations of tissue dissociation and cell lysis.

In this study, we benchmarked CosMx WTX against scRNA-seq. By processing adjacent sections from the same FFPE block tissue on both platforms, CosMx WTX demonstrated a highly comparable detection efficiency (e.g., median ~1,200 transcripts and 780 genes per cell in colon cancer) with scRNA-seq, delivering consistent cell identification for seven major cell types. CosMx WTX produced data on over 95% of the cells in the sample, without dissociative loss of smooth muscle and goblet cells in colon tissues typically seen in scRNA-seq. A similar trend was observed in other tissue types.

CosMx WTX high-throughput and unbiased nature enabled the detection of rare cell types and states, including tumor-infiltrating lymphocytes, largely missed in our scRNA-seq data due to sample size limitations and droplet variability. Despite equivalent panel size (18,935 genes in CosMx WTX vs. 18,082 genes in scRNA-seq), the spatial context enables analyses like pathway activity and ligand-receptor interactions, ensuring cellular signaling and molecular reactome are accurately mapped in their native tissue architecture. This results in data aligned with true biological processes.

Our cross-platform evaluation highlights CosMx WTX accuracy, scalability, and versatility, positioning it as a transformative tool for various research applications. Its unparalleled spatial resolution and whole transcriptome coverage set it apart, offering a more comprehensive and spatially informed view of cellular function and tissue structure. As a result, CosMx WTX has the potential to address traditional scRNA-seq applications, while simultaneously enabling biological insights with spatial context.

Accelerating microbiome meta-analysis and biomarker discovery with Cosmos-Hub's no-code platform

Manoj Dadlani[1], Kelly Moffat[1], Rita Colwell[1,2]

1 CosmosID, Inc., Integrative Systems Biology, Germantown, MD
2 University of Maryland Institute for Advanced Computer Studies (UMIACS), College Park, MD

The rapid expansion of microbiome research has resulted in an abundance of data, creating challenges in integrating, standardizing, and analyzing findings across diverse studies. We introduce the Cosmos-Hub, a user-friendly software platform designed to streamline meta-analysis of large-scale microbiome datasets without the need for coding expertise. Leveraging advanced microbial profiling algorithms, the Hub offers ultra-high-resolution taxonomic, functional, phage, anti-microbial resistance, and virulence detection, providing detailed insights into microbial communities.

In a case study, we conducted a meta-analysis of over 1,000 samples from individuals with inflammatory bowel disease (IBD) and healthy controls. Utilizing integrated statistical tools such as Maaslin 2 and Maaslin 3, we identified significant differences in microbial composition, functional pathways, and phage populations associated with IBD. This comprehensive approach not only validated known biomarkers but also uncovered novel candidates, highlighting new possibilities for drug development and diagnostic applications.

A standout feature of the Cosmos-Hub is allowing researchers to seamlessly incorporate their own sequencing data and compare it with over 60,000 microbiome profiles available in our expansive Data Warehouse. This functionality enables users to contextualize their findings within a broader spectrum of existing data, enriching the depth of their analyses. Additionally, the platform's SRA import functionality easily allows the inclusion of microbiome data from any publication, further augmenting its value in accurate and rapid large-scale data analysis.

By eliminating the need for programming skills, the Cosmos-Hub democratizes advanced microbiome analysis, making it accessible to researchers across various disciplines. The platform simplifies data integration, meta-data management, high-resolution profiling, functional and phage analysis, and interpretation—all through an intuitive interface, with no coding skills required. By accelerating the pace of data-driven discoveries, the Cosmos-Hub has the potential to significantly advance the field of microbiome research.

Compression sequencing for deep and scalable single-cell multi-omics

Mingjie Dai[1],[2], Yan Yan[1]

1 Rice University, Department of Bioengineering, Houston, TX
2 Rice University, Department of Chemistry, Houston, TX

Single-cell and spatial transcriptomics have broadly transformed biological research and biomedicine. However, current single-cell studies are fundamentally limited by sequencing depth at both molecular and cellular levels, and face significant challenges in mRNA detection sensitivity and cell profiling throughput. Consequently, faithful detection of low-abundance transcripts or analysis of rare cell types in complex sample remain challenging and costly. Whereas recent combinatorial and emulsion-based approaches make highly scalable cell barcoding possible (100,000~1M cells), sequencing cost scales linearly with total number of cells, and poses significant limitation on the scale of current studies, making large atlas efforts limited to a few well-funded labs, and high-throughput drug screens or large cohort studies inaccessible.

Current single-cell sequencing approaches are fundamentally limited by the high dynamic range of mRNA expression in human cells (5-6 logs). Specifically, >95% of all sequencing reads are taken by a small fraction of high-abundance genes (<10%) – highly redundant and contribute little to cell clustering or gene expression analysis. On the other side, complete coverage of low-abundance genes require ~1M reads per cell, which is 50-100x higher than a standard scRNA-seq experiment. A method that effectively suppresses the "redundant reads" from high-abundance genes while preserves the "critical reads" from low-abundance genes, will allow significant savings on total sequencing depth, as well as allow deep sequencing of low-abundance gene targets, and highly scalable single-cell profiling.

We developed a new biochemical strategy, "compression sequencing", that effectively compresses the dynamic range in scRNA-seq samples, thus reduces redundant reads and allows for deep and scalable single-cell sequencing. Specifically, we developed a PCR-based molecular workflow, that allows unbiased and quantitative compression of gene abundances in highly multiplexed context. We demonstrated our method in scRNA-seq sample (barcoded with 10X 3' gene expression kit), and showed complete cell coverage, higher (2X) mRNA detection sensitivity, 1,000X reduction of gene dynamic range, and overall 20-50X reduced sequencing depth and cost. Our method is broadly applicable for sequencing-based multi-omic methods, and promise to enable deep and scalable single-cell and spatial multi-omics in the future.

High-throughput profiling of cell-to-cell interactions using Nanovial Multicell Assays

Joseph de Rutte[1], Shreya Udani[1], Ronald Rondon[1], Katie Kershaw[1], Jesse Liang[1]

1 Partillion Bioscience, Los Angeles, CA

Understanding cell-to-cell interactions and their functional outputs at the single-cell level is critical for advancing immunology, oncology, and cell therapy development. Cellular secretions play a central role in this communication, orchestrating immune responses, tumor microenvironments, and regenerative processes. However, current technologies lack the resolution and throughput required to profile these interactions at scale.

Here, we introduce Nanovial Multicell Assays, a high-throughput platform designed to address these challenges using commercially available flow cytometers and single-cell sequencing workflows. This system enables the co-localization and functional profiling of interacting cell pairs in nanoliter-scale compartments, allowing the analysis of millions of cells simultaneously. Using hydrogel-based Nanovials, we encapsulate multiple cell types to preserve cell viability and function. This platform allows for the measurement of secreted proteins—including antibodies, cytokines, and growth factors—as well as cellular responses from neighboring cells, such as binding, activation, and gene expression.

As a proof of concept, we demonstrate this technique's utility in antibody discovery applications. By optimizing cell capture antibodies, ⊠-IgG secretion detection cocktails, and assay conditions, we achieve efficient co-localization of antibody-secreting (Human, Murine, CHO) and antigen-expressing (Human, CHO) cells. The platform achieves high-throughput analysis through flow sorting, typically capturing over 50% of antibody-secreting cells paired with antigen-expressing cells in individual Nanovials. We further illustrate the assay's power using a model system of ⊠-human CD3 IgG-secreting cells and CD3-expressing Jurkat cells, identifying agonistic ⊠-CD3 antibodies based on changes in Jurkat cell gene expression.

Our results showcase how Nanovial Multicell Assays overcome key challenges in antibody discovery, particularly in bispecific antibody screening and cell-reporter assays. By maintaining cell viability post-screening, the platform enables further downstream characterization. This approach provides unprecedented insights into cell-to-cell dynamics, offering a powerful tool for drug discovery and cellular analysis.

Molecular Inversion Probes (MIPs) represent a valuable approach for the monitoring and conservation of the endangered Marsican Brown bear

Massimo Delledonne[1],[2], Marzia Rossato[1],[2], Luca De Antoni[1], Leonardo Vincenzi[1], Riccardo Focaia[1], Alfonso Esposito[1], Patrizia Giangregorio[3], Nadia Mucci[3]

1 University of Verona, Department of Biotechnology, Verona, Italy
2 Genartis srl, Verona, Italy
3 Italian Institute for Environmental Protection and Research (ISPRA), Department for the Monitoring and Protection of the Environment and for Biodiversity Conservation, Ozzano dell'Emilia (Bologna), Italy

Monitoring wild populations using genetic markers is a powerful approach in conservation biology, allowing the assessment of genetic diversity and population structure. Microsatellites are traditionally analyzed for this aim. However, they may not be sufficiently variable to uniquely distinguish different individuals, especially in small (highly inbred) populations. In addition, their analysis can be technically demanding, requiring high-quality DNA and precise laboratory techniques to avoid errors in allele scoring, a critical aspect when using environmental specimens. Microsatellite investigation can also be prone to technical issues such as stutter bands and allele dropout, which can affect the accuracy of genotyping. Single-nucleotide polymorphisms (SNPs) offer several advantages, including higher resolution in detecting fine-scale population structure and subtle genetic differences, allowing for precise estimates of population-level diversity and differentiation. The analysis of polymorphic SNPs with massive parallel sequencing allows the efficient investigation of mixed samples, from non-invasive samples like hair or feces, with the possibility to evaluate Variant Allele Frequency (VAF) precisely. In addition, SNPs can help to identify genetic variations associated with local adaptation, providing insights into how populations adapt to their environments. In this work, we adopted Molecular Inversion Probes (MIPs) combined with Illumina sequencing for SNP analysis to monitor the Marsican Brown bear, a critically endangered species of center Italy consisting of ~50-60 remaining individuals. The MIP technology allowed the capture of 106 SNPs, known to be polymorphic in the bear population, by employing redundancy (4 distinct MIPs for each SNP), thus minimizing the risk of allele drop-out. We selected a 75bp fill-in distance between probe pairs to allow the analysis of degraded DNA samples. MIP probes contained 5nt-Unique Molecular Identifiers (UMI) for the precise estimation of VAF in heterogeneous/mixed samples. Genotyping analysis based on MIPs showed robust performances, with a very high SNP-calling rate (96% on average) and concordant calling in invasive (blood or tissues) and the respective non-invasive specimens (hairs), the latter characterized by high DNA degradation and low concentration.

NOTE – NON CONSIDERARE

(continue to page 146)

(continued from page 145)

ALTERNATIVA: As such the approach is expected to have a reduced allelic drop-out and improve the accuracy of monitoring the population of Marsican Brown bear at conservation aims.

The MIP panel was successfully applied on 35 DNA from invasive (blood or tissues) or non-invasive specimens (hairs), showing high degradation and low concentration. At least 92 SNP were genotyped per sample and genotype calling was concordant in the invasive and respective not-invasive samples. The generated data are being analyzed to define population structure and familial relationships.

Monitoring small wild populations using genetic markers is a powerful approach in conservation biology and it is traditionally performed using microsatellites analysis. Although microsatellites are useful for assessing genetic diversity and population structure, they also suffer some limitations. In small populations, they may not sufficiently variable to uniquely distinguish different individuals.

Monitoring small wild populations using genetic markers is a powerful approach in conservation biology, allowing assessing genetic diversity and population structure. Microsatellites are traditionally analyzed at this aim, however they suffer some limitations. In small populations, microsatellites may be not sufficiently variable to uniquely distinguish different individuals. Ultimately their analysis can be technically demanding, requiring high-quality DNA and precise laboratory techniques to avoid errors in allele scoring, a critical aspect when using non-invasive specimens, such as hairs and feces.

This redundancy mitigates the potential impacts of new or uncharacterized variants that may lead to uneven coverage or allelic dropout in PCR-and hybrid capture-based methods

Unique molecular barcodes for use in low-input and high-sensitivity applications such as variant calling from cf/ctDNA.

Molecular Inversion Probes (MIPs) represents a valuable approach for the monitoring and conservation of wild species

An improved TempO-LINC single-cell assay sensitively resolves cellular heterogeneity from multiple tissues and sample types

Dennis Eastburn[1], Nathan Jayne[1], Kevin White[1], Salvatore Camiolo[1], Gioele Montis[1], Seungeun Ha[1], Joanne Yeakley[1], Bruce Seligmann[1]

1 BioSpyder Technologies, Inc., Carlsbad, CA

We previously reported on the development and basic performance of TempO-LINC, a novel, targeted and probe-based genomics platform for single-cell transcriptome profiling that is derived from and uses the same commercial assay content as TempO-Seq. TempO-LINC combines the benefits of the TempO-Seq assay, including the lack of a reverse transcription step, with a combinatorial split pool barcoding approach to provide transcriptome-wide gene expression measurements at single-cell resolution, identifying up to 100,000 fixed cells or nuclei per run.

We now report on an improved version of this assay and its application to profiling nuclei from three distinct mouse tissues, cerebral cortex, kidney and lung. Unsupervised clustering of single nuclei run with the TempO-LINC assay was able to resolve the complex diversity and representative distributions of cells from these tissues, resulting in the identification of over 50 unique cell types. For each tissue type dataset, we further analyzed the cluster-specific expression patterns of canonical cell type marker genes to further support the clustering and annotation of our single-cell data. Over 130 biomarker genes in total were analyzed and dot plotted across the 50 annotated cell clusters. The data also demonstrates the sensitive detection of rare cell types such as pulmonary neuroendocrine cells and kidney podocytes, with only 7 cells of the latter being found within 10,794 total cells or 0.06% of the population. Additionally, we demonstrate that the targeted nature of the TempO-LINC approach as well as the ability to quantitatively attenuate select genes resulted in a low percentage of cell-associated reads occupied by transcripts from mitochondrial, ribosomal and highly expressed housekeeping genes. This effectively increases the sensitivity/gene detection rate of TempO-LINC and likelihood that the approach will identify more informative genes expressed at lower levels.

Lastly, we will report on current studies using TempO-LINC to uniquely address two challenging single-cell genomics applications, the study of alternative splicing in drug treated breast cancer cells and transcriptome profiling of archived FFPE samples from multiple human cancers. In both examples we show how the probe-based TempO-LINC assay affords increased sensitivity relative to scRNA-Seq approaches that rely on reverse transcription. In sum, our results indicate that TempO-LINC can generate rich datasets from complex tissues and can be applied to investigate cellular heterogeneity within samples that were not previously amenable to single-cell sequencing approaches.

A novel liquid biopsy-based approach enables the diagnosis and management of hematological malignancies

Yael Michaeli[1], Noa Gilat[1], Topaz Kreiser Shem Tov[1], Gal Mesika Surizon[1], Orit Feldstein[1], Mika Golan[1], Mika Bell[1], Neryia Halevi Fessler[1], Miri Neaman[2], Jasline Deek[2], Yuval Ebenstein[2], Shahar Zirkin[1]

1 JaxBio Technologies LTD., Netanya, Israel
2 School of Chemistry, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel

The diagnosis and monitoring of hematological malignancies currently depend on invasive procedures like bone marrow aspiration or lymph node biopsy. Continuous invasive tests are vital for evaluating disease progression and detecting minimal residual disease (MRD). Moreover, these invasive tests are critical for monitoring pre-malignant conditions that may develop into acute malignancies, allowing early intervention.

To address these clinical challenges, JaxBio introduces a minimally invasive blood test for the screening and early diagnosis of hematological malignancies. This approach targets epigenetic biomarkers in DNA from both peripheral blood cells and cell-free DNA. The process uses chemoenzymatic reactions to fluorescently label epigenetic marks, and a custom microarray chip to identify disease-specific signatures. Funded by the EU, this test is being developed under the SANGUINE project, with a multi-center prospective clinical trial involving 3,000 patients currently conducted across Europe.

Here, we present interim results from the SANGUINE trial, covering nine hematological malignancies and pre-malignant conditions. A detailed analysis of epigenetic marks in more than 400 cell-free DNA and whole blood samples revealed distinct signatures for each condition, composed of 50-100 biomarkers. This biomarker panel forms the basis for the custom "HemaChip," which provides a low-cost, simple, and ultra-sensitive method for cancer detection.

Our findings show that the test can correctly classify lymphoma as well as multiple myeloma with >90% accuracy and acute myeloid leukemia (AML) with 100% accuracy. Notably, the biomarker panel distinguishes follicular lymphoma (FL) from diffuse large B-cell lymphoma (DLBCL), offering the potential to detect the early transformation of FL to DLBCL. Additionally, the panel differentiates AML from the pre-malignant condition myelodysplastic syndrome (MDS) and multiple myeloma (MM) from its pre-malignant states, MGUS and SMM, enabling better disease monitoring and early intervention.

These results represent a significant breakthrough in hematological malignancy diagnostics and monitoring. The "HemaChip" offers an accurate, non-invasive alternative for detecting hematological cancers and pre-malignant conditions, reducing the need for invasive procedures and facilitating earlier detection.

Nanowire-assisted, single-molecule sequencing by synthesis with standard industry sequencers

Søren M. Echwald[1], Eric Yu, Niklas Mårtensson[1], Karin Blume[1], Kenneth W. Harlow[1]

1 Aligned Bio AB, Lund, Skaane, Sweden

The ability to detect single molecules of DNA during the sequencing process in sequencing by synthesis (SBS) type chemistries circumvents the need to amplify the sequence prior to performing sequencing. Single-molecule sensitivity thus eliminates one of the cardinal drawbacks of SBS chemistry: phasing errors. Phasing errors have limited SBS-based sequencing to reads of 150-200 bps before error rates increase to levels that make single bp changes difficult to detect. The sequencing market has been seeking low-error, cheap, medium to long-read sequencing for years in order to address a number of diagnostic questions connected with single-base pair changes in sequences ranging from 400-3000 bps. The ability to convert standard, SBS chemistry-based, short-read sequencers (e.g., Illumina MiSeq) to long-read capability would extend the utility of these instruments for sequencing applications requiring highly accurate sequencing in the range of 400-2000 bps. Aligned Bio's nanowire-based sensor platform technology provides enhanced detection of fluorescent molecules and is being adapted to allow cost-effective, plug-and-play flow cell modifications that convert these types of instruments to long-read sequencers. In addition to supporting greatly increased read length of these standard instruments, the elimination of the need for an amplification step in the sequencing process will bring the added benefits of eliminating the introduction of sequence errors related to the amplification process as well as streamlining the workflow and increasing throughput. This should lead to an overall increased efficiency and cost-effectiveness of nanowire-enhanced, single-molecule SBS over traditional methods of SBS sequencing.

Automating PacBio library construction: scaling the Darwin Tree of Life project at the Sanger Institute

Ben Farr[1], Eric Yu, Shelly-Ann Coutts[1], Lesley Shirley[1], James Watts[1], Craig Corton[1], James Uphill[1], Ian Johnston[1], Paul Lomax[2], Maryia Karpiyevich[2], Gregory Young[3], Camille Conner[3]

1 The Wellcome Sanger Institute, Hinxton, Cambridge, United Kingdom
2 SPT Labtech, Melbourne Science Park, Melbourne, United Kingdom
3 Pacific Biosciences, Menlo Park, CA

The Wellcome Sanger Institute (WSI) is a world leader in genome research that delivers insights into human, evolutionary and pathogen biology. To deliver the Institute's scientific goals the WSI generates high-quality data through the translation of ambitious experimental ideas into high-throughput automated pipelines.

A flagship programme is the Darwin Tree of Life (DToL), which aims to generate platinum standard reference genomes of the estimated 70,000 eukaryotic species in Britain and Ireland. This initiative aims to advance our understanding of life on Earth by unravelling the genetic blueprints of a wide range of organisms.

To address the scale of this challenge, the WSI is evaluating robust automated solutions to enable the generation of long-read next-generation libraries. Here we describe the development and validation of a novel automated method on SPT Labtech's firefly® liquid handling platform to construct whole genome sequencing (WGS) libraries from genomic DNA using the PacBio SMRTbell prep kit 3.0 (SPK3). Libraries were subsequently sequenced on PacBio's fleet of long-read systems. We will show how we established that firefly was a viable platform, and will demonstrate how the prep on firefly performed relative to a non-automated prep when processing DNA from a test panel composed of 8 different phylogenetically diverse species.

The best of both worlds: affordable analysis of difficult regions with short- and long-read hybrid pipelines

Technology and Application Development

POSTER 427

Donald Freed[1], Hong Chen[1], Zhipan Li[1], Brendan Gallagher[1], Luoqi Chen[1]

1Sentieon Inc., San Jose, CA

Genetic testing laboratories are increasingly considering WGS as a first-line diagnostic test, providing enhanced coverage of non-exonic regions and better detection of CNV, SV, and repeat expansions than WES. PacBio HiFi long-read sequencing offers a compelling alternative, with higher variant calling accuracy in difficult-to-map regions, better SV, methylation, and phasing detection, and enhanced resolution of medically important genes like SMN1, GBA, and PMS2. Despite these advantages, the higher cost and lower throughput of long-read sequencing instruments are a barrier to broader adoption.

Here we present a new hybrid secondary analysis pipeline that takes as input full-coverage short-read sequencing and low-coverage PacBio HiFi sequencing. By leveraging long-read haplotypes to help improve short-read alignment, our method produces more accurate results than either technology alone at 30x coverage.

We benchmark the performance of the hybrid pipeline across a variety of datasets including the Genome in a Bottle (GIAB) draft Q100 and the v4.2.1 truthsets. The new pipeline with 30x NovaSeq and 10x HiFi coverage makes 5,909/1,665 (SNP/indel) errors relative to the GIAB v4.2.1 HG002 truthset. The pipeline reduces variant calling errors by 74% (23,462/5,366 errors) compared to DeepVariant with 30x NovaSeq data and by 26% (3,226/7,023 errors) compared to DeepVariant with 30x HiFi data. We see similar improvements across all other GIAB samples, and when using the draft Q100 truthset. The new pipeline has a unique advantage across long homopolymers and tandem repeats, where it reaches higher accuracies than can be achieved by existing single technology pipelines. We demonstrate that the hybrid pipeline identifies medically relevant variants in SMN1 that are difficult or impossible to resolve from short-reads alone. While we focus here on human data, this method also works with non-human species. Overall, the pipeline demonstrates resolution of difficult genes that is comparable to or better than full-coverage long-read approaches with inputs that are a fraction of the cost.

Optimization of single-cell sequencing from archival tissues and comparison of short-read sequencing platforms

Elizabeth A.R. Garfinkle[1], Corinne Strawser[1], Jaye B. Navarro[1], Jocelyn Lucyshyn[1], Anthony R. Miller[1], Sophia Paxton[2], Jesse Westfall[1], Adithe Rivaldi[1], Patricia McCallinhart[3], Aaron J. Trask[3,4], Adrienne Dorrance[5], Derek Warner[6], Katherine E. Miller[1,7], Elaine R. Mardis[1,4,7]

1 Nationwide Children's Hospital, Institute for Genomic Medicine, Columbus, OH
2 The Ohio State University, Biomedical Sciences Graduate Program, Columbus, OH
3 Nationwide Children's Hospital, Center for Cardiovascular Research, Columbus, OH
4 The Ohio State University College of Medicine, Department of Pediatrics, Columbus, OH
5 Univeristy of Utah, Huntsman Cancer Institute, Division of Oncology, Salt Lake City, UT
6 University of Utah, DNA Sequencing Core, Salt Lake City, UT
7 The Ohio State University College of Medicine, Department of Neurosurgery, Columbus, OH

The range of available short-read sequencing platforms and compatible input material for single cell RNA-sequencing (scRNA-seq) has grown in recent years. We report a comparison of sc-RNA-seq across several short-read sequencing platforms including the Illumina NovaSeq6000 and NovaSeqX+, Singular Genomics G4, and the Element AVITI. We sequenced four single cell libraries (two human and two mouse) on the different sequencers and found near perfect correlations of single cell clusters and gene expression. In addition to scRNA-seq from fresh/frozen tissue and single cell suspensions, it is now feasible to extract whole cells and nuclei for scRNA-seq from archival samples such as those stored in formalin-fixed paraffin-embedded (FFPE) and optimal cutting temperature (OCT) compound blocks. This approach opens the door for retrospective and correlative single cell transcriptomic profiling of banked patient samples and serial timepoints over the course of a patient's treatment. We have optimized scRNA-seq from FFPE and compared the data to traditional single nuclei RNA-sequencing (snRNA-seq) from paired frozen tissues. We also directly compared scRNA-seq from FFPE and OCT blocks. First, we compared two human epileptic brain samples for which we had paired frozen tissue and FFPE blocks. Although similar numbers of total single cells were sequenced, we found a high degree of variability in the types of cell clusters identified and, therefore, a very low correlation of gene expression. This is likely due to differences in cell composition between frozen tissue sections and the corresponding FFPE blocks and highlights considerations that need to be taken if merging data from different tissue storage conditions. In contrast, when we compared scRNA-seq on cells extracted from paired FFPE and OCT blocks from two mouse heart samples, we found highly correlative cell clusters and gene expression profiles, likely due to the more uniform cell populations across the samples and storage conditions. Collectively, these results demonstrate scRNA-seq samples sequenced on different instruments can faithfully be merged for downstream analysis. We also show that scRNA-seq can be performed from FFPE and OCT blocks; however, differences in tissue section composition needs to be considered when comparing with single cell data derived from fresh or frozen samples.

Targeted transcriptome sequencing of up to one million immune cells allows identification of rare cell types at low cost

Gary Geiss[1], Simone Marrujo[1], Alexa Suyama[2], Sarah Schroeder[1], Crina Curca[1], Bryan Hariadi[1], Joey Pangallo[1], Efthymia Papalexi[1], Gokhan Demirkan[1], Charles Roco[1], Alexander Rosenberg[1]

1 Parse Biosciences, Inc., Research and Development, Seattle, WA
2 Oregon Health Sciences University, Portland, OR

Since its inception, single-cell RNA-sequencing (scRNA-seq) has been applied across many cell types and multiple fields of research, leading to novel biological discoveries. The number of cells per single cell experiment has grown exponentially with combinatorial barcoding, enabling profiling of millions of cells per project.

An approach to reduce the sequencing per cell would enable even larger projects with more cells. Specific applications of scRNA-seq, such as cell type identification or gene pathway analysis, often do not require sequencing of the entire transcriptome, but rather use specific sets of well-characterized genes. In these cases, sequencing the entire transcriptome may be adding unnecessary sequencing reads, especially when projects scale to hundreds of thousands or millions of cells.

Here, we show how enriching 100s to 1000s of genes of interest can reduce sequencing needs by >10 fold. To illustrate the power of our technology, we demonstrate the enrichment of canonical immune cell markers and pathways in whole transcriptome libraries of one million peripheral blood mononuclear cells (PBMCs) from twelve donors as well as 730,000 human bone marrow mononuclear cells (BMMCs) from four leukemia and four healthy donors. This method increased the percent of reads on target from 8% in the whole transcriptome libraries to 60% in the PMBC enriched libraries while maintaining the cell type proportions found in the whole transcriptome. Moreover, our targeted sequencing method enabled the detection of a rare (~0.35%), leukemia-specific T cell population that showed highly proliferative and cytotoxic potential.

Overall, we demonstrate that this modular enrichment strategy preserves biological structure, allows for deep characterization of rare disease-relevant cell types, and enables scaling the number of cells with less sequencing.

Assessing HiFi genome sequencing as first-tier test in rare disease genetics

Christian Gilissen[1], Ronny Derks[1], Lot Snijders Blok[1], Amber den Ouden, Simone van den Heuvel[1], Michael Kwint[1], Jeroen van Reeuwijk[1], Raoul Timmermans[1], Jordi Corominas Galbany[1], Bart van der Sanden[1], Lydia Sagath[1], Tom Hofste[1], Helger Yntema[1], Arthur van den Wijngaard[3], Janneke Weiss[1], Lisenka Vissers[1], Alexander Hoischen[1,2]

1 Radboud University Medical Center, Department of Human Genetics and Radboudumc Research Institute for Medical Innovation, Nijmegen, the Netherlands
2 Radboud University Medical Center, Department of Internal Medicine, and Radboud Expertise Center for Immunodeficiency and Autoinflammation and Radboud Center for Infectious Disease (RCI), Nijmegen, the Netherlands
3 Maastricht University, Department of Clinical Genetics, Maastricht, the Netherlands

Long-read sequencing (LRS) technologies, such as PacBio HiFi genome sequencing on Revio, allow to assess the full human genome for the first time and have the potential to revolutionize human genetics. This technology could offer a comprehensive first-tier test for clinical genetics and rare disease research.

To determine the clinical utility of LRS, we performed Revio HiFi genome sequencing on 500 human genomes with ~30-fold coverage. This included 100 mutation-positive controls with variants that are impossible or difficult to detect by standard short-read genome sequencing and necessitate a broad range of complementary test modalities in diagnostic laboratories. In addition, we sequenced 100 severe sporadic rare disease cases as patient-parent trios, and 100 severe rare disease cases as singletons who remained undiagnosed after standard-of-care testing.

Based on the 100 mutation positive samples, we find that relevant variant callers readily re-identify the majority of variants (120/145, 83%), including ~90% of structural variants, SNVs/InDels in homologous sequences and expansions of short tandem repeats. Another 10% (n=14), including aberrant methylation, skewed X-inactivation, and regions of homozygousity, was visually apparent in the data but not automatically detected. Our analyses also identified systematic challenges for the remaining 7% (n=11) of variants such as the detection of AG-rich repeat expansions. Titration analysis showed that 90% of all automatically called variants could also be identified using 15-fold coverage.

Analysis of 100 undiagnosed sporadic cases showed that de novo mutations can be readily identified on all variant levels (SNV, InDel, SV), and first cases have been diagnosed with previously hidden or missed (de novo) mutations. Preliminary data suggest >10% extra yield in undiagnosed trios and singletons, however systematic interpretation is currently ongoing.

Overall, the results of the first 500 Revio LRS genomes highlight the comprehensive nature of latest genome sequencing methods and supports LRS as a first-tier test for clinical genetics and rare disease research in the near future.

High-throughput proteomic profiling powered by NGS reveals biological insights into the tissue specificity of disease in 164 cancer cell lines

Raffaella Giugliano[1], Ryan Lamers[1], Klev Diamanti[1], Kirsty Wienand[2], Andrew Tuttle[2], Simon Zhang[3], Fransisca Vazquez[2], Catarina Campbell[3]

1 Olink Proteomics AB (part of Thermo Fisher Scientific), Uppsala, Sweden
2 Cancer Dependency Group, Broad Institute of MIT and Harvard, Cambridge, MA
3 Cancer Data Science, Broad Institute of MIT and Harvard, Cambridge, MA

Olink® technology coupled with next-generation sequencing (NGS) is the cornerstone of next-generation proteomics. Its industry-leading specificity provides accurate protein measurements that promote an in-depth understanding of biology in clinical applications, shaping the future of translational medicine.

Cultured cells are excellent biological models for studying the tissue specificity of diseases and provide greater insights into subsequent drug targets. This study leveraged the Dependency Map (DepMap) portal to perform broad proteomic profiling of 164 cell culture lysate samples, covering 45 cancers and 22 distinct tissues. The study aimed to provide further insights into cancer drug development.

Proteomic profiling of cell lysates using Olink® Explore HT (targeting >5,400 proteins) revealed that cellular protein profiles accurately reflected tissue and disease-specific biology. Multi-dimensional statistical analyses showed that protein profiles clustered together according to tissue types. For example, two types of brain tumors (neuroblastoma and glioma) formed distinct clusters from each other and exhibited discernible differences. Moreover, the hierarchical clustering of cellular protein profiles revealed unique protein signatures for each tissue and cancer type, with each forming separate clusters.

Pathway enrichment analysis highlighted unique tissue-specific patterns. Protein profiles of cultured brain cells were enriched in neuronal system pathways, while cultured cells representing lymphoid tissue were enriched in immune system pathways. Thus, the protein profiles obtained with Olink Explore HT accurately reflected the underlying biology across tissues, diseases, and pathways.

Subsequent analyses compared which proteins were detected in cell lysates and plasma based on subcellular location. Four-fold more “extracellular” or “secreted” proteins were detected in plasma than in cell lysates, whereas 25-fold more “intracellular” and “nuclear” proteins were detected in cell lysates than plasma. Thus, Olink Explore HT effectively measured the target proteins in cell lysates when their concentrations were within the detectable range.

Olink Explore HT is a powerful tool for revealing biological insights across different sample matrices. This high-throughput proteomics platform supports the understanding of disease tissue-specificity and can be used to investigate the tissue of origin for metastasis.

Exploring tumor microenvironment heterogeneity through multiomic in situ profiling of RNA and proteins

Veronica Gonzalez Muñoz[1], Thanutra Zhang[1], Smritee Dadhwal[1], Anatalia Robles[1], Cheyenne Christopherson[1], Kristen Pham[1], Joshua Thao[1], Minkyung Park[1], Hsin Wen Chang[1], Janine Hensel[1], Cedric Espenel[1], Hayato Ikoma[1], Gabriele Girelli[1], Paul Ryvkin[1]

1 10x Genomics, Pleasanton, CA

Multiomic detection of RNA and protein targets within the same tissue section provides a comprehensive view of molecular interactions and spatial organization. This integrated approach enhances understanding of cellular processes, signaling pathways, and molecular networks, leading to improved tissue analysis and disease characterization. Notably, tumor characterization and immune phenotyping are critical areas where multiomic analysis has the potential to reveal insights into tumor heterogeneity, immune cell infiltration, and treatment responses.

Here, we introduce the ability to simultaneously detect up to 28 protein markers alongside 480 genes on the Xenium in situ platform. We developed a protein panel targeting key Immuno-Oncology markers that enables thorough interrogation of the tumor microenvironment (TME) with essential markers to identify immune cells present in the tissue and their phenotypically distinct states; resolves the tissue architecture and origin of the tumor with markers such as CD31, b-catenin, vimentin, e-cadherin, and aSMA; and identifies the checkpoint mechanisms present in the TME with markers including PD-1/PD-L1, LAG-3, and VISTA.

We analyzed various healthy and cancerous human FFPE tissues using this combined RNA and protein profiling method. While most proteins corresponded with transcript levels of their coding genes, some markers showed differences in RNA and protein abundance. We focused on B-catenin, encoded by the CTNNB1 gene, which is involved in normal cellular processes and tumorigenesis. Literature suggests a poor correlation between gene and protein expression, despite a link between gene expression and disease progression. Subcellular localization, regulated post-translationally, is also indicative of disease progression. We observed discrepancies between gene and protein expression across tissues, including liver, skin, and breast. Protein staining enabled subcellular localization analysis, offering insights into B-catenin's role in tumorigenesis.

These observations emphasize the need for multiomic integration, particularly in cases of discordant expression patterns. The Xenium platform's ability to integrate high-resolution spatial gene expression with protein profiling, combined with post-run H&E imaging, provides a powerful approach to understanding disease mechanisms.

Immuno-oncology analysis of kidney, lung, colon, and breast FFPE tissue in-situ on the G4X Spatial Sequencer

Kenneth Gouin III[1], Michael Lawson1, Andrew Pawlowski[1], Sabrina Shore[1], Yuji Ishitsuka[1], Daan Witters[1], Eli N. Glezer[1]

1 Singular Genomics, San Diego, CA

Molecular level spatial profiling of tissue promises to yield new insights into tumor biology and immunotherapeutic response. Our novel platform, the G4X Spatial Sequencer, enables clinical scale studies, with simultaneous detection of RNA transcripts, proteins and H&E staining in FFPE tissues. With up to 40 cm2 of tissue area imaged in a single run, it expands throughput by an order of magnitude over existing methods.

We designed ~300-gene transcript panels for each of four tissue types: kidney, lung, colon, and breast, with an emphasis on immuno-oncology relevant targets. The panels share a common set of 150 immuno-oncology and 50 stromal genes. Each panel also includes a set of ~100 tissue-specific genes to profile cell types, cancer pathways and tumor microenvironment.

We imaged gene and protein expression in 5-µm thick sections of human kidney, lung, colon, and breast, from both healthy and cancer FFPE tissue samples. Using a novel tissue transfer process, we arrayed multiple tissue sections in two formats: a large sample format (10 mm x 10 mm) with 10 sections/slide, and a small sample format (4.5 mm x 4.5 mm) with 32 sections/slide. Each slide was analyzed in three modalities (RNA, proteins, fluorescent H&E). Following antibody staining, probe binding and amplification of targets, images were acquired using 4-color SBS readout, followed by nuclear and cytoplasmic staining to generate a fluorescence-based H&E stain.

In FFPE cancer samples and healthy controls, we detected an average of 50-200 transcripts/cell and 20-50 unique genes/cell, with the dynamic range extending to > 500 transcripts/cell. Using 6 rounds of priming, we achieved high transcript density profiling while maintaining a false discovery rate ~1%. Strong correlation between single-cell transcript and protein expression profiles confirmed specificity of each modality. Comparison across serial sections showed strong concordance in expression profiles for both modalities, demonstrating robust reproducibility. Finally, we identified distinct tumor-associated cellular phenotypes and spatial neighborhoods using a novel multi-modal graph-based clustering method leveraging RNA, protein, morphology, and spatial information.

Enspyre: a novel enrichment technology for selected DNA variants using pyrophosphorolysis

Katarzyna Anton[1], Timon Heide[1], Paulina Powalowska-Pickton[1], Ernesto Lowy-Gallego[1], Amy Lovell[1], Jeffrey Gregg[1], Sophie Hackinger[1], Magdalena Stolarek-Januszkiewicz[1], Robert J. Osborne[1], Barnaby W. Balmforth[1]

1Biofidelity Ltd., Cambridge, United Kingdom

Enspyre (Enrichment by selective pyrophosphorolysis and release) is a novel target enrichment technology developed at Biofidelity with a broad spectrum of potential applications that allows for selective enrichment of specified variants prior to sequencing. The selective enrichment with Enspyre is dependent on pyrophosphorolysis (PPL), an exquisitely specific exonuclease reaction mediated by a DNA polymerase which digests probes with full complementarity to targets but is halted by a mismatch. Consequently, only variants perfectly matched to their probes are released after probe digestion, phase-separated from the background and sequenced.

We present here proof of concept data of enrichment by Enspyre compared to a non-selective assay. A machine learning-based probe design algorithm developed at Biofidelity is used to design selective probes for Enspyre without the need for testing or optimization. 98% of probes designed with this algorithm have a selectivity score above 10, meaning mutant to wildtype molecule ratios following the assay increased at least 10-fold. Using these probes, Enspyre detects single variants with a limit of detection (95%) of 0.5%, sensitivity comparable to other NGS-based variant detection assays, but from only 5 million reads per sample. Enspyre can also be used to predict the input VAFs for variants. One of the potential practical uses of Enspyre is minimal residual disease (MRD)-tracking in oncology. In an MRD proof of concept using 1800 variant-specific probes and custom-made contrived samples, Enspyre accurately detected the presence of MRD down to 10 parts per million (ppm) with only 5 million reads per sample.

In summary, Enspyre can identify variants with a sequencing depth reduced by 95% compared to a non-selective assay. Additional benefits include lower data processing time and storage. For clinics, Enspyre could bring down the cost of sequencing necessary for MRD detection, while a shorter time of data processing would result in a quicker diagnosis for patients.

Achieving ultimate accuracy with nanopore sequencing data for cancer surveillance

Technology and Application Development

POSTER 435

Courtney Hall[1], Jared Simpson[2,3], Winston Timp[1,4]

1 Johns Hopkins University, Department of Biomedical Engineering, Baltimore, MD
2 Ontario Institute for Cancer Research, Toronto, Ontario, Canada
3 University of Toronto, Department of Molecular Genetics, Toronto, Ontario, Canada
4 Johns Hopkins University, Department of Molecular Biology & Genetics, Baltimore, MD

Single-molecule sequencing from Oxford Nanopore Technologies (ONT) offers the potential to generate real-time results on the small, affordable MinION device, making it a promising option for rapid and cost-effective cancer surveillance. Long ONT reads have proven essential for resolving challenging genomic regions and detecting complex structural variants, but their per-base accuracy has historically lagged behind sequencing-by-synthesis (SBS) technologies. This is especially challenging in situations where each read must be precisely measured, such as detecting low-frequency mutations in circulating cell-free DNA (cfDNA). Incorporating unique molecular identifiers (UMIs), which link PCR-amplified reads back to their original templates for error correction, reduces background noise in ONT reads and enables the accurate detection of variants associated with measurable/minimal residual disease (MRD) and acquired treatment resistance in cfDNA. Here we explore the ultimate accuracy of nanopore sequencing using UMI-based approaches and assess its potential in cancer surveillance.

We first show that UMI incorporation and consensus polishing can generate higher single-molecule accuracy than previously thought possible on ONT sequencing platforms. Regions of interest were tagged with dual UMIs, amplified, and sequenced on MinION flow cells. Using the "super" accurate basecalling model, reads were binned by UMI pair, and consensus sequences were generated with multiple rounds of polishing. This approach reduced the error rate of raw ONT reads over 100-fold, achieving accuracies above 99.998%.

We then applied our methods to capture low-frequency mutations from small amounts of short cfDNA from patient samples. Initial results suggest that this approach significantly reduces background noise in nanopore reads, enabling the sensitive detection of rare variants across both MRD and acquired treatment-resistant panels. With further development of our methods and continued advancements from ONT, we anticipate that UMI nanopore sequencing could provide a rapid, cost-effective solution for cancer diagnostics and monitoring at the point of care.

Significant optimization of sequencing performance for hard-to-read genome sequences, including improvements in long homopolymers via modification of DNA clustering reaction conditions

Carolyn G. Conant[1], Camila Columbo[2], Philip Balding[2], Jason Betley[2], Christopher Edlund[1], Ilaria Grizzi[2], Alix Kwasniewska[2], James Han[1], Amy Charles[2], Panos Petsagkourakis[2], Alex Calderwood[2], Tian Yuan[2], Chris Winnard[2], Cande Rogert[1], Mark Wang[1], Emily Parker[1]

1 Illumina Inc., Core R&D, San Diego, CA
2 Illumina Inc., Core R&D, Cambridge, Cambridgeshire, United Kingdom

Repetitive sequences create challenges for alignment and variant calling, because of low complexity, heterogeneity, movement, and duplication in the human genome and are becoming especially important in the era of telomere-to-telomere data, where these sequences are found abundantly. Sequencing artifacts in these loci are marked by a higher rate of mismatches and increased soft-clipping and may impact variant calling negatively. To optimize performance in these challenging regions, the reaction conditions for DNA cluster generation were modified and assessed with PCR-free whole-genome sequencing using the benchmarking genome NA24385 (HG002) in XLEAP-SBS™ chemistry. This modified reaction was evaluated on NextSeq™ 2000 P3 and P4 XLEAP-SBS configurations with DRAGEN™ 4.3.6 (graph aligner) and demonstrates significant error reduction and improvements in missed calls associated with strings of repeated nucleotides and dinucleotide motifs. On average, an 8% reduction in small variants errors was observed, with sustained improved performance in low-complexity regions. Notably, the impact of improved resolution of homopolymers and flanks is only partially demonstratable using the NIST 4.2.1 benchmark, as it only covers only 69% of these regions. For longer homopolymers, which are particularly challenging to sequence, the modified reaction conditions yielded a significant improvement in accuracy compared to the standard conditions.

Analytical validation of MRD profiling with tumor-informed whole-genome-ppmSeq

Technology and Application Development

POSTER 437

Youngmok Jung[1], Stephanie Ferguson[1], Sangmoon Lee[1], Brian Baek-Lok Oh[1], Ariel Jaimovich[2], Seungmin Nam[1], Dawoon Jung[1], Young Seok Ju[1]

1 Inocras Inc., San Diego, CA
2 Ultima Genomics Inc., Fremont, CA

Minimal residual disease (MRD) detection is critical for managing cancer, allowing for early identification of relapse and guiding treatment decisions. Most existing MRD assays rely on deep targeted sequencing of a limited number of somatic mutations, with a limit of detection (LOD) in the range of about 100–1000 parts-per-million (PPM). In this study, we present analytical validation of an innovative approach that combines CancerVision™ (Inocras) with ppmSeq™ (Ultima Genomics), integrating tumor whole-genome sequencing (WGS) with whole-genome circulating cell-free DNA (cfDNA) sequencing. By capturing a comprehensive range of cancer-related mutations and using an ultra-sensitive sequencing method, our approach significantly enhances the sensitivity of MRD detection, achieving an LOD of 5–20 PPM (5×10^{-6} to 2×10^{-5}).

We conducted an analytical validation study using three tumor and matched-normal cell line pairs: HCC2218, HCC1395, and NCI-H2126 from the American Type Culture Collection. Tumor DNAs were diluted into matched-normal DNAs at concentrations ranging from 10^{-2} to 10^{-7}, simulating various levels of circulating tumor DNA (ctDNA). These samples were sequenced using ppmSeq™ at a depth of 29–55x, yielding a mixed (duplex) rate of 25–39% and an absolute sequencing error rate of 6.3×10^{-7}. By integrating 51,350–169,079 cancer-specific somatic mutations from CancerVision™ with advanced bioinformatics algorithms to remove sequencing artifacts, our method successfully detected tumor DNA at dilutions as low as 3 PPM (3×10^{-6}). This represents a significant improvement over traditional targeted sequencing approaches in detecting MRD.

This whole-genome tumor-informed cfDNA MRD sequencing method offers a next-generation solution for non-invasive cancer monitoring. Whole genome ppmSeq™ also provides a quick turnaround and straightforward assay that does not require design and storage of bespoke panels. It holds promise for earlier relapse detection, real-time tumor dynamics tracking, and more precise therapeutic decisions, potentially leading to better outcomes for cancer patients.

A novel on-flow cell library prep for both high-quality short reads and proximity information in one prep that delivers an expanded genome

Fiona Kaper[1], Jacqui Weir[2], Mitch Bekritsky[3], Qing Zhang[2], Vicki Thomson[2], Ines Vitoriano[2], Arun Subramaniyan[1], Michael Ruehle[3], Drew Ellershaw[2], Shunhua Han[1], Daniel Andrews[2], Stephen Gaffney[2], Ritu Kundu[2], Gavin Parnaby[1], Pascal Grobecker[2], Maria Gardzielewska[2], Ivana Armogida[2], Maya Bajracharya[2], Ningxin Ouyang[1], Louise Fraser[2], Cande Rogert[1], Rami Mehio[1]

1Illumina Inc., San Diego, CA
2Illumina Inc., Cambridge, Cambridgeshire, United Kingdom
3Illumina Inc., Remote, United States

High quality short read whole genome sequencing (WGS) delivers exceptional quality throughout most of the genome, but still struggles with more challenging regions and variant types. These challenges include variant calling in difficult-to-map regions such as pseudogenes and segmental duplications, calling structural variants – especially with sequence resolution, and variant phasing.

Here, we present a highly simplified on-flow cell library preparation method on the NovaSeq™ X which delivers an expanded genome by enhancing short read sequencing data with proximity information. We describe how proximity information is leveraged to recover those genomic regions and variant types that are especially challenging in standard short read WGS and how these insights deliver an expanded genome. We also demonstrate how input DNA quality impacts proximity metrics and the resulting ability to phase across whole genes.

To determine performance of this novel on flow cell method, the proximity analysis has been demonstrated on a range of cell lines and blood samples with a range of DNA qualities. We present improvements in small variant calling performance across the genome and highlight specific examples in difficult-to-map medically relevant genes e.g., SMN1/2 and PMS2. We also describe improvements to phasing haplotypes, enabling >98% of genes to be fully phased in a standard GM24385 (HG002) cell line sample and demonstrate how fully phased genes resolve compound heterozygotes in CFTR. Finally, we describe novel methods to enable the detection of large variants, repeat expansions and complex rearrangements with the proximity mapping of short reads. We demonstrate improvements across the genome with specific examples, such as repeat expansion in DMPK and FXN and previously characterized balanced structural variants.

In Situ detection of viral HPV expression in human cervical cancer with Xenium Prime 5K

Stephanie Kravitz[1], Juan Pablo Romero[1], Ian Fiddes[1], Amanda Janesick[1], Payam Shahi[1], Sarah Taylor[1]

1 10X Genomics, Pleasanton, CA.

Spatial transcriptomics has emerged as a powerful tool to explore the spatial organization of cell types, interactions, and cell states by quantifying and localizing transcripts within intact tissues. The Xenium Analyzer provides an end-to-end solution for spatial analysis with highly sensitive and specific detection of RNA, enabling high-resolution exploration of the underlying transcriptional landscape.

Here, we demonstrate in situ detection and subcellular localization of over 5,000 genes in human cervical cancer FFPE tissue using the Xenium Prime 5K Human Pan Tissue and Pathways panel. Additionally, we designed a custom panel of 100 genes tailored to interrogate cervical cancer biology, including Human Papillomavirus (HPV) probes, to study viral influence on tumor biology.

We integrated this ultra high-plex + custom gene panel with membrane-based cell segmentation, post-Xenium H&E and immunofluorescence (IF) staining, and probe-based single cell gene expression (Flex) to generate a comprehensive, multimodal dataset in HPV+ cervical cancer tissue. This approach enhanced cellular phenotyping and the characterization of distinct niches within the cancer tissue that corresponded to HPV +/- tumor cell subtypes, spatial localization of divergent cell states and signaling throughout the tumor. The combination of spatial RNA detection with H&E and IF staining within the same tissue section provides complementary morphological insights by connecting gene expression with conventional pathology methods to interrogate HPV-driven cancer pathology.

The inclusion of HPV-16 and HPV-18 probes revealed high specificity for HPV-16, confirming the patient's infection with this subtype. Leveraging the viral RNA signal, we categorized tumor cells based on viral RNA content, allowing us to characterize the cellular composition surrounding infected and non-infected tumor subtypes. Our analysis revealed a reduced proportion of T cells adjacent to HPV-infected tumor cells. Furthermore, ligand-receptor analysis highlighted increased activity in angiogenesis-related pathways, suggesting a link between viral infection and tumor vascularization.

This study demonstrates the power of incorporating custom probes targeting non-endogenous viral RNA with spatial transcriptomics and multi-modal tissue analysis to gain comprehensive insights into the cellular and spatial dynamics of HPV+ cervical cancer.

Methods for sequencing cell-free DNA and cell-free RNA from human plasma

Karen McKay[1], Margaret Heider[1], Chen Song[1], Jian Sun[1], Gautam Naishadham[1], Kaylinnette Pinet[1], V K Chaithanya Ponnaluri[1], Bradley W. Langhorst[1], Keerthana Krishnan[1]

1 New England Biolabs, Ipswich, MA, USA

Cell-free DNA (cfDNA) and cell-free RNA (cfRNA) are nucleic acid fragments that freely circulate in blood serum. Cell-free nucleic acids (cfNAs) can originate from various types of cells, including normal cells, cancer cells, and fetal cells. A great deal of biological information can be gathered from the sequencing of cfNAs. Thus, cfNA sequencing can serve as a non-invasive approach for the monitoring of atypical and typical biological processes ranging from cancer to ageing. The ability to analyze cfDNA and cfRNA from blood samples facilitates longitudinal studies and enables researchers to track molecular changes over time via liquid biopsies.

Here we demonstrate methods for sequencing cfDNA and cfRNA isolated from healthy donor plasma samples using the NEBNext UltraExpress DNA and RNA library prep kits. Library preparation and sequencing protocols were optimized to accommodate the size and quality of the cfNAs. Using these methods, we obtained consistent, high-quality sequencing metrics including high library yield, even coverage, and high library complexity from varied individual samples. These libraries are compatible with whole genome sequencing (WGS), whole transcriptome, and target enrichment workflows. These robust and streamlined protocols for cfDNA and cfRNA sequencing can be applied to a broad range of research and diagnostic scenarios, enabling scientists interested in the biological insights available from cell-free nucleic acids.

In-situ sequencing of T-cell and B-cell receptors at sub-cellular resolution in cells and tissue

Tung T. Le[1], Zane Hiatt[1], Ryan Shultzaberger[1], Kenneth Gouin III[1], Michael Lawson[1], Daan Witters[1], Eli N. Glezer[1]

1 Singular Genomics, San Diego CA

In situ sequencing of variable gene regions is a key unmet need in understanding biology and pathology, which is of interest for applications such as clonotyping immune cells in tissues. Here we present Direct-Seq, a novel technology using the G4X platform for in situ RNA sequencing at sub-cellular resolution, demonstrated by sequencing IgH and TCRB receptors in PBMCs and FFPE tissue.

We developed a probe panel that targets the conserved V and J-regions that flank CDR3 on IgH transcripts and a second panel that targets the V and C-regions that flank CDR3 and J regions on TCRB transcripts. PBMCs were seeded onto a glass substrate, followed by fixation. The diverse gap sequence was reverse transcribed into the probes, which were then clonally amplified. Clonal amplification products were sequenced in situ using Singular Genomics' SBS chemistry. We also applied the technique to tissue samples by transferring 5-µm thick FFPE tonsil sections to G4X slides, followed by deparaffinization and antigen retrieval. Amplicon preparation and sequencing were conducted in the same way as for PBMCs.

Direct-Seq successfully profiled ~ 60% of all B-cells (relative to CD19-positive fraction) and ~ 20% of all T-cells (relative to CD3-positive fraction) in PBMC populations. We observed the expected high diversity of clonotypes, with 92-98% of B and T-cells having distinct CDR3 sequences. Comparing multiple 50 bp reads within the same cell, we calculate the sequencing accuracy to be ~ 98%. For cells that had at least 1 read, we detected on average ~ 6 reads/cell for B-cells and ~ 2 reads/cell for T-cells. For FFPE tonsil samples we observed similar levels of diversity (94-99%) and accuracy (97-99%) as we did in PBMCs, but the detection efficiency was significantly lower, likely due to reduced RNA accessibility and degradation resulting from fixation.

In conclusion, we have demonstrated the novel capability of Direct-Seq to perform in situ sequencing of IgH and TCRB transcripts in PBMCs and FFPE tissues. The ability to map the spatial distribution of immune cell clonality and diversity will offer a powerful tool to study immunity and disease.

Development of a high-throughput multiplexed NGS library preparation workflow with self-normalizing adapters

Kristin Butcher[1], Owen Smith[1], Sean Tighe[1], Tong Liu[1], Tiffany Truong[1], Cibelle Nassif[1], Danny Antaki[1], Esteban Toro, Elian Lee[1], Ramsey Zeitoun[1], Siyuan Chen[1]

1 Twist Bioscience, Research and Development, South San Francisco, CA

High throughput next-generation sequencing (NGS) workflows are required for population-level genomic research that aids in the investigation and diagnosis of genetic disorders. The cost and effort required in processing individual samples remain a barrier to efficient NGS library preparation. We present several innovative and automation-friendly technologies that reduce cost and time: Normalization by Ligation (NBL) and library preparation multiplexing with inline barcodes. We present data from experiments performed using our high-throughput automated workflow on 384 real-life human samples.

Enzymatic fragmentation (EF) of samples is the foundation of our ultra-high-throughput method. The EF module allows researchers to adjust the insert size based on sequencing format and permits using samples of varying quality without significant base composition bias. Normalization using NBL produces equivalent coverage for samples of 30 ng to 300 ng DNA input, eliminating the need for initial input quantification. NBL integrates inline barcodes to permit multiplexing post-ligation. This generates samples that can be flexibly grouped in multiples of 12 prior to clean-up and amplification. As a result, the reaction footprint and reagent usage of downstream steps decrease by 12-fold, increasing throughput and generating libraries in less time. By controlling conversion from each sample, NBL enables plexing up to 96 in the target enrichment step versus the 12-plex commonly used in other protocols. When paired with automated methods, sample capacity is greatly increased with minimal hands-on time needed.

Utilizing a novel NBL technology, inline barcodes, and automation for NGS library preparation boosts productivity and decreases hands-on time. These workflow advancements maintain uniform conversion to enable low-pass whole genome sequencing and when paired with capture, provide sufficient coverage for genotyping applications (>20x). This workflow significantly advances the experimental process of library preparation to complement throughput advancements made in sequencing instrumentation.

Decoding T Cell Activation: Spatial and Quantitative Analysis of Membrane Proteins using Molecular Pixelation

Louise Leijonancker[1], Hanna van Ooijen[1], Vincent van Hoef[1]

1 Pixelgen Technologies AB, Stockholm, Sweden

Membrane protein organization plays a crucial role in T cell function, yet traditional protein analysis methods, such as flow cytometry and CITE-Seq, fail to capture spatial information about their targets. On the other hand, fluorescence microscopy, while being the standard for spatial investigations, is limited by its multiplexing capability and throughput. In this study, we applied Molecular Pixelation (MPXTM), a novel sequencing-based technology for high-multiplex spatial analysis of membrane proteins on single cells.

MPX enables simultaneous mapping of 80 proteins on 1000 individual cells per sample, allowing for a detailed comparison of the membrane proteome of cell populations, here applied to compare resting and activated T cells. Our results highlight pronounced differences in protein abundance of more than 25 proteins, including upregulation of key activation markers such as CD69, CD25, and CD38.

In addition, we could reveal dramatic spatial protein re-organization, exemplified by the polarized distribution of CD82 in a subset of activated T cells. These findings emphasize the importance of spatial proteomics in understanding T cell activation and suggest that MPX holds great potential as a discovery tool for unravelling immune cell surface architecture and function.

Methods for Transcriptomic Profiling of Cerebrospinal Fluid to Identify Biomarkers in Schizophrenia

Joshua Z. Levin[1], Jixiang Zhang[1], Sean K. Simmons[1], Matthew B. Johnson[1]

1 Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

Biomarkers are important tools for disease diagnosis and designing effective clinical trials, particularly for brain disorders such as schizophrenia that are highly heterogenous. Cerebrospinal fluid (CSF) analysis can aid in diagnosis of diseases and conditions affecting the brain and spinal cord. The Schizophrenia Spectrum Biomarkers Consortium (SSBC) is creating a biorepository containing CSF and other specimens from individuals with schizophrenia spectrum disorders and controls. Although the primary focus of SSBC is to test specific hypotheses about potential disease protein biomarkers, our study is designed to extend SSBC by profiling RNA in CSF to generate new hypotheses about schizophrenia biomarkers. As such we have developed a comprehensive methodology for profiling RNA in CSF samples. This approach addresses the inherent challenges of low RNA input in CSF cell-free supernatants and low cell numbers in pellets.

We have demonstrated feasibility for RNA profiling of CSF supernatants with bulk RNA-seq and cell pellets with single-cell RNA-seq (scRNA-seq). For supernatants, we will report comparative results for multiple RNA isolation and library preparation methods starting from as little as 10 pg RNA, which is equivalent to 100 µL of CSF supernatant. We produced robust results – detecting >8,000 genes in a reproducible manner. Most of the genes detected are consistent with gene expression found in immune cells, though there is also expression of brain-enriched genes. For cell pellets, we implemented a protocol for the thawing and recovery of cells (Beth Stevens lab, personal communication), maintaining a cell viability rate of 90%. We successfully profiled the RNA, including T cell receptor (TCR) and B cell receptor expression, in these cells. Our analysis revealed over 5,000 cells recovered from 10 mL of CSF and identified all the expected cell types. The vast majority of cells expressing TCRs were T cells, indicating the efficacy of the TCR labelling process and the reliability of the experimental results. We also compared our CSF scRNA-seq data to that of peripheral blood mononuclear cells (PBMCs) and found similar cell types in PBMCs and CSF, though there are large differences in proportion. Future work will focus on profiling CSF samples from SSBC and other sources.

Automated high-plex protein detection in cells for mapping drug response in a model system

Tyler Lopez[1], Pin Ren[1], Jennifer Wong[1], Ben Song[1], Daniel Honigfort[1], Mariam Dawood[1], Adeline Mah[1], Sanchari Saha[1], Colin Brown-Greaves[1], Derek Fuller[1], Michael Kim[1], Grace Yeo[1], Tristin Rammel[1], Marielle Krivit[1], Jane Yao[1], Ramreddy Tippana[1], Vivian Dien[1], Connor Thompson[1], Shuyang Ye[1], Ryan Kelley[1], Matthew Kellinger[1], Amirali Kia[1], Sinan Arslan[1], Michael Previte[1]

1 Element Biosciences, inc, Research and Development, San Diego, CA

Protein detection in cell-based assays is crucial for understanding cellular responses to various stimuli. Current methods, including immunofluorescence (IF), mass spectrometry (MS), and western blotting, each offer unique advantages but come with inherent limitations. High-plex protein detection with IF provides single-cell resolution but is constrained by host organism compatibility and extended workflows, typically detecting only 1-10 proteins per assay with turnaround times of 24-72 hours. In contrast, MS can identify thousands of proteins in a single run but lacks single-cell resolution and demands extensive technical expertise, with turnaround times ranging from 3 to 7 days. To overcome these challenges, we developed a scalable technique utilizing sequencing by avidity to achieve high-plex protein detection with single-cell resolution, delivering results within 24 hours. This innovative approach allows for deeper insights into the complex signaling networks involved in cellular responses. One key area of interest is the role of tumor necrosis factor-alpha (TNFa), a pro-inflammatory cytokine that induces apoptosis in various cell types, including HeLa cells, through its receptors TNFR1 and TNFR2. The p38 MAPK inhibitor SB202190 modulates this apoptotic response by inhibiting the p38 pathway, enhancing TNFa-induced apoptosis and preventing survival signals that counteract TNFa's effects. By sensitizing cancer cells to TNFa, SB202190 amplifies the apoptotic signal, highlighting the potential for targeted cancer therapies. In this study, HeLa cells, a human cervical cancer model, were treated with TNFa (100 ng/mL) to induce apoptosis. Inhibitors were introduced to disrupt TNFa's standard signaling pathway, prompting cells to rely on alternative mechanisms for survival and proliferation. Cells were fixed, permeabilized, and analyzed using Aviti24, which allows simultaneous profiling of more than 100 proteins involved in apoptosis, MAPK, and cell cycle regulation, offering single-cell resolution in a 12-well format. Aviti24 successfully identified distinct activation pathways and cellular escape mechanisms in HeLa cells treated with TNFa and its inhibitors. Detailed pathway mapping revealed how inhibitors targeting specific nodes in the apoptotic and MAPK pathways redirected cellular signaling, uncovering survival strategies and potential therapeutic targets. Aviti24 enhances the monitoring of protein dynamics with single-cell resolution. By offering scalability, rapid 24-hour results, and comprehensive pathway coverage, Aviti24 provides valuable insights into cellular behavior under pharmacological perturbation and holds promise for advancing cancer treatment strategies.

InterlACE: A multiomic platform for simultaneous profiling of unmethylated DNA and genomic variants for improved therapy selection and patient stratification

Debora Lucarelli[1], Luca Tosti[1], Calum Mould[1], Valentina Miano[1], Anthony C. Smith[1], Jack Kennefick[1], Robert K. Neely[1,2]

1 Tagomics Ltd, Suite 3 Cori Building, Granta Park. Great Abington, Cambridge, United Kingdom, CB21 6DF
2 The University of Birmingham, School of Chemistry, Edgbaston, Birmingham, B15 2TT, UK

Multiomic DNA analysis offers the potential for a more comprehensive insight into cellular function and disease states. However, scaling this approach has been challenging due to the complexity of aligning multiple workflows and integrating the diverse, heterogeneous data types.

Tagomics' InterlACE platform directly addresses these challenges, providing a fully integrated solution for multiomic DNA profiling. The platform incorporates their proprietary epigenetic, conversion-free, profiling technology, ActiveACE, which enables highly efficient genome-wide enrichment of unmethylated CpG sites1, with combined whole-genome or targeted sequencing in a single streamlined workflow.

Here, we introduce the integration of Agilent Technologies' SureSelect Library Preparation and Cancer CGP Assay into the InterlACE multiomic workflow. The SureSelect targeted panel allows for the detection of key somatic variants in solid tumours, including single nucleotide variants (SNVs), copy number variants (CNVs), indels, translocations, gene fusions, and immuno-oncology biomarkers such as tumour mutational burden (TMB) and microsatellite instability (MSI). The InterlACE platform not only retains the sensitive variant-calling capability of SureSelect but also adds Tagomics' genome-wide epigenomic profiling, delivering both genetic and epigenetic insights from the same input sample.

By integrating genetic and epigenetic profiles, InterlACE has the potential to revolutionise personalised medicine and accelerate drug development, offering rich biological data to support target discovery and patient-specific insights to aid in stratification.

We present a case study demonstrating how InterlACE was used to generate personalised multiomic profiles for tumour samples, resulting in deeper biological insights into the patient's underlying disease state and exposing novel markers that could further enhance the patient's stratification and treatment options.

References

Luca Tosti, Calum Mould, Imogen Gatehouse, Anyela Camargo, Anthony C. Smith, Krystian Ubych, Peter W. Laird, Jack Kennefick and Robert K. Neely, Epigenomic profiling of active regulatory elements by enrichment of unmodified CpG dinucleotides, bioRxiv, 2024.02.16.575381

MeCUT&RUN: An Ultrasensitive, Low-Cost Approach for Mapping DNA Methylation

Technology and Application Development

POSTER 447

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

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

DNA methylation (DNAme) is one of the most widely studied epigenetic mechanisms of gene regulation and is associated with a broad range of aging and age-related diseases, including neurodegeneration and cancer. Understanding how DNAme signaling drives physiological dysfunction has great potential to reveal novel biomarkers for therapeutic strategies. While several approaches have been developed to map DNAme genome wide, these assays are severely limited for clinical research due to high cell input and sequencing requirements and/or issues with sequence bias (e.g., due to DNA fragmentation steps or antibody bias). This includes the current gold standard, whole genome bisulfite sequencing (WGBS), which uses a destructive fragmentation workflow and requires high cell inputs (>1M cells) and sequencing depth (>400M reads).

To address these limitations, we are developing MeCUT&RUN, a modular platform that provides an ultrasensitive and high-throughput approach for profiling DNAme genome wide. This technology utilizes novel recombinant Methylation Sensors (MeSensors) as highly specific detection reagents for DNAme in an immunotethering strategy inspired by CUT&RUN. In the MeCUT&RUN workflow, cells or nuclei are incubated with a MeSensor to tether protein A/G-micrococcal nuclease (pAG-MNase) for targeted cleavage. The clipped methylated DNA (meDNA) is then processed using one of two downstream library preparation and sequencing workflows. For researchers seeking the most efficient approach, the enriched DNA may be directly sequenced to produce a broad map of DNAme (MeCUT&RUN-Fast) with nucleosome-level resolution (~150 bp). Alternatively, researchers can enzymatically convert unmethylated cytosines to achieve base pair resolution (MeCUT&RUN-EM).

To develop this assay, we first engineered a novel MeSensor and validated its specificity using methylated nucleosome substrates. We then applied the MeSensor to develop both MeCUT&RUN-Fast and MeCUT&RUN-EM workflows. We compared MeCUT&RUN-EM directly to whole genome enzymatic methyl-seq (EM-seq) and observed that >80% of methylated CpGs detected by EM-seq were also captured using MeCUT&RUN-EM, but at a fraction of the cost due to reduced sequencing needs. Overall, we demonstrate that MeCUT&RUN is highly versatile, sensitive, and requires 10-30x fewer sequencing reads than existing approaches. This technology will make DNAme profiling more widely accessible and cost-effective, empowering novel studies into DNAme mechanisms, biomarkers, and drug development.

Application of PacBio Reduced Sample Input Workflow in Pediatric Genomic Profiling

Anthony R. Miller[1], Huachun Zhong[1], Maria E Hernandez Gonzalez[1], Alejandro Otero Bravo[1], Benjamin J Kelly[1], Don Corsmeier[1], Jesse Hunter[1,2], Melanie Babcock[1,2,3], Catherine E Cottrell[1,2,3], Elaine R Mardis[1,2,4], Richard K Wilson[1,2]

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

Genomic profiling through sequencing has significantly advanced the detection of genetic variations linked to disease, particularly in pediatric and adolescent young adult populations. Long-read sequencing technologies, such as Pacific Biosciences (PacBio) HiFi reads, offer key advantages. These include the ability to span repetitive regions, resolve complex genomic areas, and characterize structural variants—crucial for studying disease-associated variants and population genetics. However, genomic profiling in pediatric patients is challenging due to the limited availability of genomic DNA (gDNA), as there are constraints on how much blood or tissue can be safely collected. Additionally, the gDNA extracted is often prioritized for clinical assays, leaving little material for research. This creates a need for workflows that require lower gDNA input.

PacBio has recently developed a new workflow for loading samples onto the Revio system which improves loading efficiency by 4x, which reduces the gDNA input required for SMRTbell construction. We demonstrated the utility of this reduced sample input workflow by sequencing gDNA from a NICU neonatal patient. Anticipating the benefits of this reduced sample input workflow, we reduced the input gDNA from 2,000 ng to 1,200 ng and incorporated stringent sizing with Sage BluePippin due to slight gDNA degradation. Final library prep resulted in a SMRTbell library with an average size of 20 kb at a concentration of 17.5 ng/µl in 15 µl—a molarity too low for standard Revio HiFi sequencing at 200 pM. Yet, the improved reduced sample input workflow allowed sequencing of up to two Revio SMRT Cells at the desired 200 pM concentration. Sequencing a single Revio SMRT Cell yielded 54% P1 loading and a mean HiFi read length of 12.4 kb, resulting in 29X coverage.

This patient was initially enrolled into our clinical rapid genome sequencing protocol, where short-read sequencing revealed an expansion in the polyA region of the PHOX2B gene. Long-read sequencing provided enhanced resolution and confidence in the called variant, demonstrating the reduced sample input workflow's ability to generate high-quality data from limited input. This is an important advancement for pediatric genomic research, now enabling HiFi sequencing for patient samples previously ineligible due to input requirements.

BGE: Development of a blended genome and exome sequencing service at the Wellcome Sanger Institute

Nicole Müller-Sienerth[1], Matthew Mayho[1], Sara Widaa[1], Jamie Lovell[1], Marcella Ferrero[1], Sandra Carrizosa Mataix[1], Richard Rance[1], Ben Farr[1], Diana Rajan[1], Mike Quail[1], Carol Scott[2], Michael Blackhurst[3], Sujit Dey[3], Katy Taylor[2], Liz Cook[1], Ben Topping[2]

1 Wellcome Sanger Institute, Sequencing Operations, Cambridge, UK
2 Wellcome Sanger Institute, Pipeline Solutions, Cambridge, UK
3 Wellcome Sanger Institute, Portfolio Delivery Team, Cambridge, UK

Understanding the genetic causes and biological mechanisms of disease susceptibility and progression requires large scale studies looking at genomic variation within diverse cohorts of people. In an ideal world the whole genome would be sequenced at a great depth, however, this is going to be prohibitively expensive! A compromise approach is to combine exome sequencing at high depth to detect rare variants with microarray genotyping to look at common genome wide variants. However, this relies on external providers and costs, turnaround times here have increased significantly and it is susceptible to some biases.

But how do we effectively capture the diversity of the genome at scale? “Blended Genome Exome” (BGE) technology offers a new solution. First developed and used at Broad Institute of MIT and Harvard, Cambridge, MA, blending an ISC exome library with a whole genome library is a good alternative to the above mentioned methods and is very cost effective. The resulting low coverage genome (2x-4x) and high coverage exome (30x-40x) data from the same sequence run (single CRAM file) can then be used to detect unbiased common variants at whole genome level as well as rare coding variants.

Adapting the BGE process as a sequencing service for researchers at the Wellcome Sanger Institute has been a good example of multi-disciplinary teamwork with staff contributing to this project in many different ways. As the main elements of the process (PCR-free whole genome sequencing and ISC exome capture) already exist within Sequencing Operations, BGE should be a simple process to implement. Here we want to highlight early successes and challenges we faced during the development of this project.

Comparison of NGS Library Preparation Methods for Enhanced Variant Detection

Technology and Application Development

POSTER 450

Michal Naftali[1], Ofer Isakov[2], Dina Yagel[1], Rotem Greenberg[2], Shay Ben Shachar[2]

1 Genomic Center, Clalit Health Services, Petah Tikva, Israel.
2 Clalit Research Institute, Clalit Health Services, Ramat Gan, Israel.

Introduction: Selecting an appropriate library preparation (LP) method is crucial for accurate detection of disease-causing variants in whole exome sequencing. Due to the large number of available LP kits, there is a need for systematic, reproducible comparison capabilities to assess performance in clinical applications accurately.

Objective: To evaluate and compare the performance of commercial LP kits for NGS applications, focusing on enrichment quality metrics and variant calling performance.

Methods: Genome in a Bottle (GIAB) reference samples were prepared using three widely-used commercial kits: 1. IDT - Nextera DNA Flex Pre-Enrichment Library Prep with xGen Exome Research Panel V2 Target Regions. 2. Illumina- DNA Prep with Exome 2.0 Plus Enrichment. 3. Twist - Exome 2.0 plus Comprehensive Exome. All preparations were performed using a Hamilton robot platform. Sequencing was performed on NovaSeq 6000 platform. Quality metrics were collected at various stages, including sequencing, alignment, and variant calling. Sensitivity, precision, and F1 scores were calculated using an intersection of the IDT and Twist target regions, and used to assess overall variant calling performance.

Results: A total of 288 samples were tested across multiple runs (96 samples for each LP process, including GIAB samples). Median coverage was 97.2[87,107], 120.9[103,134] and 99.0[78,111] for IDT, Illumina, and Twist, respectively. Percent above 20X were 95.8 and 95.5 for IDT and Illumina, respectively, while for Twist it was 97.8. Uniformity values were 95.9, 94.4 and 97.9 for IDT, Illumina and Twist, respectively. Reviewing variant calling performance using the GIAB samples, recall for SNVs were 0.986692, 0.989365 and 0.991204 for IDT, Illumina, and twist, respectively. Presicion for SNVs were 0.997084, 0.99805 and 0.99815 for IDT, Illumina, and twist, respectively. Similar trends with slightly decreased performance were observed for indels.

Conclusion: Across key metrics, about enrichment and coverage the Twist LP kit demonstrated superior performance in several key metrics. Overall, this comprehensive comparison provides valuable insights for laboratories optimizing NGS workflows for clinical applications.

Assessing genomic cell heterogeneity with semi-permeable capsules

Juozas Nainys[1], Andrius Šinkūnas[1], Jokūbas Tamoliūnas[1], Greta Rukšnaitytė[1], Paulius Matulis[1], Vaidotas Kiseliovas[1], Rapolas Žilionis[1]

1 Atrandi Biosciences, Vilnius, Lithuania

Cancer evolution and progression are driven by the accumulation of somatic mutations that confer fitness advantages to tumor cells. Tumor heterogeneity, in turn, plays a critical role in variations in treatment response. To better understand cell-lineage heterogeneity and detect rare clones, single-cell genomic analysis is an invaluable tool that retains essential information on mutation co-occurrence. Current approaches to single-cell whole-genome sequencing involve isolating cells into multi-well plates, followed by whole-genome amplification and library preparation. While whole genome coverage, scalability, single-nucleotide resolution, and cost are key factors in selecting a sample preparation protocol, existing methods fail to excel in all of these areas simultaneously.

Here, we demonstrate a scalable single-cell whole-genome sequencing workflow using the Semi-Permeable Capsule (SPC) technology. Once individual cells are encapsulated in 60 µm SPCs, the semi-permeable nature of these miniature compartments enables efficient multi-step processing and unique labeling of gDNA. This miniaturized assay enhances whole-genome amplification, reducing bias in genome coverage compared to reactions performed in multi-well plates. For proof-of-concept, we barcoded 10,000 individual cells across multiple acute myeloid leukemia patient samples. We next performed low-pass WGS for CNV analysis, as well as target gene pull-down followed by deeper sequencing from the same cells. This approach allowed us to generate a matched dataset with single-nucleotide resolution, demonstrating the scalability and precision of our method in simultaneously capturing large-scale genomic alterations and fine-resolution point mutations.

Upscale SPLAT: a high-throughput barcoding and sequencing strategy for population genomics and planetary biodiversity

Nordlund, J[1,3], Daniels RJ[2,3], Lundmark Anders[1,3], Wiman AC[1,3], Ramsell J[1,3], Pettersson M[2,3], Andersson L[2,3], Raine A[1,3]

1 Uppsala University, Department of Medical Sciences, Uppsala, Sweden
2 Uppsala University, Department of Medical Biochemistry and Microbiology, Uppsala, Sweden
3 Uppsala University, Science for Life Laboratory, Uppsala Sweden

As the consumable costs for high-throughput sequencing continue to decline, library preparation remains a disproportionately large expense, particularly for projects involving small to mid-sized genomes. This is especially problematic in population genomics and biodiversity studies, where sequencing many hundreds of samples is essential for uncovering genetic diversity, evolutionary patterns, and environmental responses. The high cost of library preparation can limit the scale of such studies, creating a bottleneck in the field.

To address this, we developed upscale Splinted Ligation Adapter Tagging (upSPLAT), a method that combines early barcoding with our single-stranded splinted ligation approach (Raine et al., NAR 2017) to enable pooled library preparation. upSPLAT supports high-plex pooling of up to 64 samples in a single library, reducing per-sample costs to ~5% of those associated with traditional single-sample library prep. Additionally, our single-strand ligation approach can handle damaged or denatured DNA, as well as viral ssDNA, making upSPLAT highly versatile for a variety of sample types.

Using upSPLAT, we prepared whole-genome sequencing libraries from ~300 DNA samples isolated from Atlantic herring and a variety of other fish species, targeting 10x coverage per individual using the Illumina NovaSeq X Plus sequencer. To assess performance, we evaluated the evenness of individual sample representation in upSPLAT pools and compared the genomic coverage across non-overlapping 10 Kb windows between a subset of upSPLAT libraries and Tn5-based libraries generated from the same DNA samples. upSPLAT demonstrated high reproducibility and accuracy across GC contents ($R > 0.7$) and consistent coverage uniformity when compared to individually prepared Tn5 libraries. upSPLAT has been successfully integrated into routine workflows at the SciLifeLab National Genomics Infrastructure in Sweden, demonstrating its scalability, reproducibility, and cost-effectiveness for large-scale population genomics and biodiversity research across diverse species with varying genome sizes and GC contents.

Rapid pathogen identification and genomic surveillance with onboard data analysis on MiSeq™ i100

Nidhi Shah[1], Tayler Mazurski[1], Priyanka Prashar[1], Jeff Koble[1], Rita Stinnett[1], Dorothea Agius[1], Elizabeth Loughran[1], Scott Kuersten[1], Emily Parker[1], Mark Wang[1], Robert Schlaberg[1]

1 Illumina Inc., San Diego, CA

Pathogen genomics has been a critical driver of discovery in the biological sciences for decades. Recently, COVID-19 has highlighted the importance of tracking viral spread and evolution in near real time. Pathogen genomic surveillance is now increasingly used in both centralized and distributed settings for pandemic preparedness, to support vaccine development and monitor effectiveness, to prevent and contain hospital-acquired and food-borne infections, identify mechanisms and spread of drug resistance, and clinical testing. Wastewater-based, systematic pathogen genomic surveillance can identify outbreaks at the community level before they become apparent as an increase in hospitalizations, serving as an early warning system. The same approach is now also being used to monitor and track the spread of antimicrobial resistance (AMR), which is threatening to reverse decades of progress in medicine. For many of these applications, rapid answers and streamlined workflows are essential.

We used three different enrichment panels (Virus Surveillance Panel V2, VSPv2; Respiratory and Urinary Pathogen ID/AMR Enrichment Panels, RPIP and UPIP) and one amplicon sequencing kit (Illumina Microbial Amplicon Prep-Influenza A/B, IMAP-Flu) to prepare libraries for targeted viral and AMR surveillance from wastewater as well as genome sequencing of influenza A and B viruses, SARS-CoV-2, and respiratory syncytial virus (RSV) from residual human samples combined with rapid sequencing on MiSeq™ i100. Resulting data was automatically analyzed onboard with the DRAGEN™ Microbial Enrichment Plus (DME+) and DRAGEN™ Targeted Microbial app (DTMA).

Multiple respiratory (incl. human adenovirus A and F, human bocavirus) and gastrointestinal viruses (incl. Aichi virus, norovirus, rotavirus, sapovirus) as well as a broad range of AMR markers were detected at fluctuating abundance in serial wastewater samples from college dorms and entire communities. RPIP enabled genomic surveillance of high priority viruses combined with broad pathogen detection using targeted metagenomics and push-button data analysis with DME+. IMAP-Flu and DTMA generally generated full-length genome sequences of influenza A (H1N1, H3N2) and B (Yamagata) viruses at >10X coverage with PCR threshold cycle values <30.

Our results demonstrate efficient pathogen genomic surveillance and broad pathogen detection in a range of sample types using off-the-shelf sequencing solutions with low input requirements, rapid (4-7h) sequencing with flexible throughput, and onboard, push-button data analysis, eliminating the need for bioinformatics expertise. Standardized and streamlined workflows could help decentralize pathogen surveillance and protect the health of individuals and communities.

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

An end-to-end enzymatic solution for methylomes from low input gDNA and cfDNA

V K Chaithanya Ponnaluri[1], Brittany Sexton[1], Vaishnavi Panchapakesa[1], Daniel J. Evanich[1], Matthew A. Campbell[1], Bradley W. Langhorst[1], Louise Williams[1]

1 New England Biolabs, Ipswich, MA

The cytosine modifications 5-methylcytosine and 5-hydroxymethylcytosine are important regulatory marks, and their identification within genomes is essential in understanding gene regulation. Methylation signatures from clinical samples like cell-free DNA (cfDNA) play a critical role in biomarker discovery and cancer diagnostics. However, the amount and quality of cfDNA available for diagnostics can be limiting. Historically, cytosine methylation has been detected using bisulfite sequencing. This method uses chemical conversion of cytosines into uracils and leads to DNA damage which results in shorter DNA insert sizes as well as biases in the data. For lower input cfDNA bisulfite conversion based methylomes are especially problematic due to the biases introduced and reduced genome coverage.

In contrast to bisulfite sequencing, NEBNext Enzymatic Methyl-seq (EM-seq), leaves DNA intact and results in superior sequencing libraries with longer insert sizes, lower duplication rates and minimal GC bias. An enhanced EM-seq workflow can now accurately detect methylation in DNA samples using as little as 0.1 ng DNA. This EM-seq v2 workflow when combined with enzymatic fragmentation (NEBNext UltraShear) enables further streamlining of library preparation by enabling end to end automation capability.

EM-seq v2 libraries were prepared using both mechanical and enzymatic (NEBNext UltraShear) fragmentation of 0.1 ng to 200 ng N12878 DNA and unfragmented 0.1 ng to 10 ng of cfDNA. Additionally, we also prepared 1 ng cfDNA libraries with both acoustic and enzymatic fragmentation. Regardless of fragmentation method, all libraries generated robust yields with even GC bias distribution, consistent insert size profile, low duplication rates and high rates of mapped reads. Furthermore, for the NA12878 libraries, ~56 million CpGs were identified for 200 ng to 1 ng inputs and ~44 million CpGs for 0.1 ng inputs. Interestingly, for cfDNA, the coupling of NEBNext UltraShear with EM-seq v2 reduced positional bias across the sequencing read, resulting in more accurate methylation calling.

EM-seq v2 combined with enzymatic fragmentation provides a robust, cost effective, end to end library preparation solution. Libraries have superior sequencing metrics that establish it as the gold standard for accurate methylation profiling, particularly for lower DNA input and challenging samples like cfDNA.

GEM-X Universal Multiplex: a cost-effective, sample and species agnostic multiplexing approach for single-cell RNA sequencing

Mohammad Rahimi[1], Ann Renschler[1], Brett Olsen[1], Daniel Reyes[1], Elliott Meer[1], Ivan Akhremitchev[1], Shamoni Maheshwari[1], Sreenath Krishnan[1], Yang Gu[1], Yarden Killmer Maya[1], Peter Smibert[1], Jessica Terry[1]

1 10x Genomics, Pleasanton, CA

In recent years, sample multiplexing has emerged as a key addition to single cell RNA sequencing workflows. It allows an increase in cell and sample throughput per experiment, improved resolution of doublets, and reduction in sample-specific batch effects. Furthermore, sample multiplexing also lowers the cost per sample

Sample multiplexing approaches typically require labeling of cells with sample-specific barcodes prior to sample pooling and/or leveraging preexisting genetic diversity to enable assignment of each cell back to its original parent sample after cell encapsulation, library preparation, and sequencing. However, these methods have several fundamental limitations, including a high cell input requirement, additional sample preparation steps, sample-specific optimization, and limited sample and/or species compatibility.

To overcome this, we present GEM-X Universal Multiplex, a cost-effective, streamlined approach to sample multiplexing. This approach simplifies multiplexing by eliminating upstream sample tagging, offers flexibility across different sample types, species types and assays, and supports lower cell inputs compared to traditional multiplexing methods.

The GEM-X Universal Multiplexing approach relies upon an easy-to-use microfluidic chip that enables the co-partitioning of cells from four samples with four unique sets of gel beads. Post encapsulation, library generation and sequencing, 10x Barcodes are used to associate cDNA-derived reads back to the individual cells, and the unique gel bead set.

We demonstrated the utility of this sample multiplexing approach on peripheral blood mononuclear cells across multiple human donors as well as multiple species and compared the data obtained between GEM-X Universal Multiplexing and GEM-X Universal Gene Expression solutions. Regardless of the approach, we obtained comparable data in terms of library quality and complexity, cell surface protein detection, the number of V(D)J clonotypes detected, and consistent cell type identification and proportions.

This new approach streamlines sample multiplexing for single cell RNA sequencing, enabling a cost effective, sample and species agnostic approach, all powered by GEM-X Universal, a consistent and reproducible platform for generating high-quality and complexity single-cell data at scale.

Dried Blood Spot gDNA extraction using the Kingfisher Apex instrument

Nicholas Redshaw[1], Iraad Bronner[1], Diana Rajan[1], Matthew Forbes[2], Katherine Figueroa[2], Cristina Ariani[2], Eleanor Drury[2], Sonia Morgado Goncalves[3], Ian Johnston[1]

1 Wellcome Sanger Institute, Sequencing Operations, Hinxton, Cambridge, United Kingdom
2 Wellcome Sanger Institute, Genomic Surveillance Unit, Hinxton, Cambridge, United Kingdom
3 Wellcome Sanger Institute, Cellular Operations, Hinxton, Cambridge, United Kingdom

The Malaria Genomic Epidemiology Network (MalariaGEN) is a network of partners from 40+ countries, which uses state-of-the-art tools and techniques to understand how genetic variation in humans, Plasmodium parasites, and Anopheles mosquitoes affects the transmission of malaria.

Dried Blood Spots (DBS) are a favoured sampling method for the MalariaGEN as they enable sequenceable material from samples to be collected in the field without the need for refrigeration, which is especially important for resource-deprived regions where the disease is endemic.

As a MalariaGEN collaborator, the Wellcome Sanger Institute has an established high-throughput, automated pipeline for the extraction of gDNA from DBS followed by selective Whole Genome Amplification, Genotyping by Amplicon Sequencing and Whole Genome Sequencing.

The discontinuation of our current instrument of choice for DBS gDNA extraction (Qiagen BioRobot Universal System) prompted the evaluation and validation of a newer instrument, the Thermo Kingfisher Apex. Here we present data from our validation experiments with the Kingfisher Apex using the MagMAX DNA Multi-Sample Ultra 2.0 Kit, and demonstrate a higher DNA extraction yield, and improved coverage and SNP calling efficiency compared to the BioRobot.

Furthermore, we discuss the numerous operational advantages of the Kingfisher Apex instrument including fast turnaround (approx. 45 min runtime for 96 samples), small lab footprint, simple interactive touchscreen control and flexibility in choice of reagents; all of which enable us to continue to support and deliver the objectives of the MalariaGEN in the coming years.

Scaling of spatial imaging transcriptomics for broad applications using 10x Genomics Xenium

Kenny Roberts[1], Katy Tudor[1], Minal Patel[1], Ian Johnston[1], Muzz Haniffa[1], Omer Bayraktar[1], Andrew Bassett[1]

1 Wellcome Sanger Institute, Hinxton, Cambridgeshire, United Kingdom

The explosion of spatial transcriptomics technologies over the past decade has revolutionised biology, providing an evermore detailed picture of cells and gene expression in biological context. A key challenge for all technologies in this field is applicability to diverse samples in a high-throughput high-reproducibility manner that generates data in a format easily interrogated by researchers and the scientific community. We have built a spatial transcriptomics pipeline around the 10x Genomics Xenium platform in order to serve the broad scientific needs of the Wellcome Sanger Institute including cell atlasing, neurodegeneration, cancer genomics, infectious disease, and evolutionary genomics.

Thus far we have processed over 750 tissue samples, mapping over 30 billion transcripts across 55 million cells, from diverse mammalian tissues, including all major human organ systems across different life stages including healthy and diseased donors, as well as mouse and marmoset. Applicability to archival (pathology) samples allows retrospective analysis of decades of diseased samples in addition to prospective samples from clinical trials, in vitro models (organoids and cells), and on-going adult and foetal tissue cohorts.

Using standardised protocols, instrumentation including cyclic automated liquid handling and imaging, and on-instrument image wrangling and analysis, our Xenium pipeline has accelerated our generation of spatially resolved transcriptomics data by at least 10- to 15-fold as compared to manual in situ sequencing in addition to allowing targeting of 3- to 5- fold more genes. Standard chemistry allows targeting of up to 480 genes and yields up to 2,000 transcripts per cell for tissue (medians up to 650) and 4,000 for cultured cells. Recently developed Xenium Prime chemistry allows targeting of over 5,000 genes, an unprecedented depth of imaging-based spatial data, and delivers up to 5,000 transcripts per cell (medians up to 1,400) in tissue. This excitingly illustrates a depth comparable to some single-cell and sequencing-based lower-resolution spatial experiments, but with sub-micron spatial resolution.

We are now further scaling our pipeline to process thousands of tissue samples with both 480-plex and 5000-plex assays, as well as diversifying targets by mapping single nucleotide variants (SNVs) to map cancer clones and perform pooled cell screens.

Platform for large-scale analysis of challenging and difficult-to-handle cell types at single-cell resolution spanning longitudinal multi-modal measurements

Gary Schroth[1], Nirvan Rouzbeh[1], Sina Farahvashi[1], Olaia Vila[1]

1 Cellanome, Foster City, CA

Single-cell technologies have revolutionized our understanding of cellular functions and heterogeneity. However, current high-throughput methods face limitations due to their reliance on single or limited modalities, cell dissociation into suspension, and flow-based microfluidics. These constraints hinder studies involving critical cell types for disease etiology. Specifically, challenges arise with: 1) cells inadequately identified by protein markers (e.g., senescent cells), 2) large or fragile cells that lose viability under harsh microfluidic conditions (e.g., neurons, myeloid cells), and 3) cells needing disruption from their functional units for transcriptional analysis (e.g., neurons). Impure identification or processes that perturb the transcriptome, especially in delicate cells like neurons, diminish the biological relevance of results.

To overcome these limitations, we developed a high-throughput, multi-modal platform facilitating the longitudinal (days to weeks) study of individual cells or groups without enzymatic perturbation. We demonstrated the platform's utility by conducting multi-modal analyses on challenging cell types, including senescent cells, neurons, fibroblasts, and monocytes at single-cell resolution. Neurons were chosen as a case study due to significant sorting challenges and the need to disrupt neuronal networks for transcriptional profiling.

Using Cellanome's platform, we captured tens of thousands of live neurons in CellCage™ Enclosures (CCEs) without enzymatic dissociation, either as isolated single cells or within functional networks. We successfully achieved: 1) long-term culture and growth of human neurons in their natural adherent state, 2) formation of functional neuronal networks validated by calcium flux assays, and 3) isolation of whole neurons or neuronal somas from interconnected neurons for single-cell multi-modal and transcriptomic analysis.

We generated single-cell RNA-seq (scRNA-seq) data from adherent neurons and observed differential gene expression compared to suspended neurons, indicating that physical context significantly influences neuronal transcriptomic states. This technology preserves cell integrity and function, unlocking new insights and providing a vital tool for studying the transcriptome and live cellular behavior.

Development of a method for targeted long-read methylation sequencing and its application to microdissected cancer specimens

Masahide Seki[1], Keisuke Kunigo[1], Satoi Nagasawa[1], Ayako Suzuki[1], Yutaka Suzuki[1]

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

Long-read sequencing technologies, such as nanopore sequencing, have enabled sequencing of long DNA molecules (>10 kb) and the detection of base modifications, including DNA methylation. The long read length facilitates the efficient analysis of repetitive regions, structural variants, and allele-specific variations that are challenging to resolve with short-read sequencing. In DNA methylation analysis, targeting specific regulatory regions at high depth is often more informative than shallow whole-genome sequencing. Although several long-read targeted DNA methylation methods have been developed, they typically require micrograms of DNA and have limited target region coverage.

Here, we present targeted nanoEM (t-nanoEM), a method that combines base conversion, nanopore sequencing, and target capture to enable high-depth, long-read DNA methylation analysis of specific genomic regions using as little as 10 ng of DNA. Using breast cancer cell lines, we demonstrated that t-nanoEM can generate reads ~5 kb in length at up to 500x depth. Furthermore, we developed a workflow to detect allele-specific methylation differences and applied it to analyze allelic methylation including imprinting regions. We applied t-nanoEM to samples microdissected from sections of breast and lung cancer tissues to perform spatially resolved DNA methylation analysis. By comparing adjacent microregions, we identified allele-specific methylation alterations associated with cancer progression. T-nanoEM has the potential to enable high-depth, long-read methylation analysis of various rare samples, including those with limited DNA availability.

Liquid Footprints: an innovative biomarker class for unbiased precision diagnostics based on the analysis of ultra-short double-stranded cell-free DNA from plasma

Kai Sohn[1], Jan Müller[1], Mirko Sonntag[1]

1 Fraunhofer IGB, Stuttgart, Baden-Wuerttemberg, Germany

Liquid Biopsy paved the way for a new era of molecular testing of complex diseases. Accordingly, specific characteristics of cell-free DNA (cfDNA) including relative abundance, mutations or epigenetic modifications are exploited for the detection of tissue- or cell type specific alterations. Consequently, circulating cell-free nucleic acids from plasma and serum provide rich information on pathologies, infectious diseases, tumor status or genetic alterations in individuals. More recently, fragmentomics of cfDNA fragments from blood plasma also gained considerable attention for applications of yet unmet diagnostic needs or performances.

We therefore established an approach to enrich and analyze for ultra-short double-stranded cfDNA fragments from human plasma. Using a size-selective electrophoretic approach for total cfDNA from human plasma we isolated fragments with an apparent molecular weight of 20-60 base pairs. Subsequent high-throughput DNA sequencing (NGS) in combination with proprietary bioinformatic workflows facilitated genome-wide DNA footprints from Liquid Biopsies ('Liquid Footprints') as enriched fragments were preferentially derived from gene promoters, binding sites of transcription factors as well as structural DNA-binding proteins. Within liquid footprint DNA from healthy individuals, we found a significant enrichment of 203 different transcription factor consensus motifs. Moreover, levels of ultra-short double-stranded cfDNA at specific genomic regions correlate with DNA-methylation and H3K4me3 histone modification levels, as well as transcription of downstream genes. When comparing pancreatic ductal adenocarcinoma with colorectal carcinoma or septic with post-operative control patient samples, we identified 731 and 1107 differentially occupied loci, respectively thus facilitating discrimination between colorectal and pancreatic cancers as well as between septic patients and clinical controls from patient plasma samples. In conclusion, analysis of ultra-short, double-stranded cfDNA fragments reveals an accurate picture of genome-wide transcription factor footprints including the occupancy of disease-specific transcription factor binding sites thus demonstrating a promising potential for diagnostic applications.

Ultra-rich single-cell targeted long-read sequencing with scTaILoR-seq

William Stephenson[1], Ashley Byrne[1], Daniel Le[1], Kostianna Sereti[2], Hari Menon[1], Samir Vaidya[1], Neha Patel[1], Jessica Lund[1], Ana Xavier-Magalhães[1], Minyi Shi[1], Yuxin Liang[1], Timothy Sterne-Weiler[2,3], Zora Modrusan[1]

1 Genentech Research and Early Development, Department of Proteomic and Genomic Technologies, South San Francisco, CA
2 Genentech Research and Early Development, Department of Discovery Oncology, South San Francisco, CA
3 Genentech Research and Early Development, Department of Oncology Bioinformatics, South San Francisco,

Single-cell RNA sequencing predominantly employs short-read sequencing to characterize cell types, states and dynamics; however, it is inadequate for comprehensive characterization of RNA isoforms. Long-read sequencing technologies enable single-cell RNA isoform detection but are challenged by limited throughput and unintended sequencing artifacts. Here we develop Single-cell Targeted Isoform Long-Read Sequencing (scTaILoR-seq), a hybridization capture method which targets over a thousand genes of interest, improving the median number of on-target transcripts per cell by 29-fold. We use scTaILoR-seq to identify and quantify RNA isoforms from ovarian cancer cell lines and dissociated tumors, yielding 10,796 single-cell transcriptomes. Targeting of T-cell genes enabled the reconstruction of immune repertoires and clonotype frequencies. Using long-read variant calling we identify a putative pathogeneic cancer mutation and reveal associations of expressed single nucleotide variants (SNVs) with alternative transcript structures. Phasing of SNVs across transcripts enabled the measurement of allelic imbalance within distinct cell populations. Overall, scTaILoR-seq is a long-read targeted RNA sequencing method and analytical framework for exploring transcriptional variation at single-cell resolution.

Utilizing ion-exchange chemistry for rapid and efficient concentration of DNA and RNA

Technology and Application Development

POSTER 464

Grittney Tam[1], Caryn Hale[1], Fanli Meng[1], Juan Blanco-Heredia[1], Sitharam Ramaswami[1], Michael F. Berger[1], Brian Loomis[1]

1 Memorial Sloan Kettering Cancer Center, Center for Molecular Oncology, New York, NY

Dilute/high volume nucleic acid samples often require concentration upstream of NGS workflows to maximize input into library prep. The most common methods for sample concentration are the use of a vacuum concentrator (if the sample is in an appropriate buffer) or the use of a SPRI-based precipitation reagent. While both methods work reasonably well, they require additional handling steps and add time to the overall process. A potentially simple, rapid and efficient alternative to these methods is an ion-exchange based approach, where nucleic acids are adsorbed onto a polycationic surface that allows for buffer exchange and volume adjustments. While ion-exchange reagents are not novel, their implementation in NGS workflows is rare, suggesting underlying issues with ease of use or recovery efficiency from commercially available reagents. To explore the use of this type of reagent in our NGS workflow, our lab built a simple ion-exchange reagent for nucleic acid concentration using inexpensive off-the-shelf reagents and carboxy-modified paramagnetic particles. Initial tests of our reagent showed quantitative recovery of input DNA from dilute solution in less than five minutes. To test the efficiency of our system in a real-world context, we concentrated cfDNA and RNA samples from dilute solutions, magnetically collected the beads, discarded the entire supernatant and put the bead-adsorbed nucleic acids directly into their relevant NGS library prep workflows. Target enrichment was performed, and the resulting pools were sequenced at appropriate depth and analyzed for library complexity, target coverage, GC bias, and noise profiles. Our data show that there are no observable differences in these metrics between control samples that were prepared without concentration and the samples that were pre-concentrated using our method. Thus, our magnetic ion-exchange system provides a rapid and simple option for lossless nucleic acid concentration steps and buffer exchanges.

Identifying protein binding partners at the bench using Quantum-Si

Winston Timp[1], Courtney Hall[1], Jessica Hosea[1], Julian Rocha[1]

1 Johns Hopkins University, Department of Biomedical Engineering, Baltimore, MD

Recent breakthroughs in DNA and RNA sequencing have expanded our understanding of genomes and transcriptomes, enabling telomere-to-telomere assemblies and profiling full-length RNA isoforms at single-cell and spatial levels. These advances were made possible in part by affordable, accessible technologies and molecular methods. In contrast, despite stunning advances in proteomics, it remains confined to specialized facilities with costly equipment, limiting accessibility and throughput. This is particularly problematic when investigating how DNA sequence changes, epigenetic states, or RNA isoforms interact with protein binding partners to regulate gene expression and phenotype. A simple, rapid method to identify these protein partners from targets is crucial.

To address this, we leveraged the Quantum-Si Platinum instrument, a benchtop device that directly sequences peptides. The Quantum-Si platform provides key insights into protein identification, relative abundance quantification, and peptide sequencing, surpassing the limitations of mass spectrometry, which relies on size/charge-based identification. Peptides bind to individual wells, where sequential recognition and cleavage of amino acids allow for precise sequencing up to 14 amino acids in length, sufficient for unique identification.

We began by sequencing standard proteins (BSA, H2A.X) and then moved on to analyzing antimicrobial resistance proteins in bacterial lysates. While we successfully sequenced peptides up to 14 amino acids, proline residues often blocked full sequence coverage. By comparing read coverage to standard curves, we quantified mixtures of antimicrobial proteins. Additionally, we applied enrichment strategies, using SDS-PAGE to isolate proteins from specific gel bands.

Next, we adapted affinity purification methods, typically paired with mass spectrometry, to investigate protein-DNA interactions. We used biotinylated nucleic acid baits to capture proteins from lysates, testing this method with lacO repeats to identify lacI. We further targeted CTCF binding motifs, comparing proteins bound to methylated versus unmethylated motifs. These methods offer new insights into how genomic and epigenomic variations affect protein binding in disease states, providing a rapid and cost effective way to identify protein partners regulated by genomic, epigenomic, or splicing variation.

Multiplexed optical barcoding and sequencing for spatial omics

Aditya Vankatramani[1], Didar Ciftci[1], Limor Cohen[1], Chris Li[1], Khanh Pham[2], Xiaowei Zhuang[1,2,3]

1 Harvard University, Department of Chemistry and Chemical Biology, Cambridge, MA
2 Harvard University, Department of Physics, Cambridge, MA
3 Harvard University, Howard Hughes Medical Institute, Cambridge, MA

Spatial omics methodologies fall broadly into two categories: sequencing-based and imaging-based approaches. While sequencing-based tools capture nucleic acids without targeting specific sequences for unbiased discoveries, their spatial resolution is often limited to larger than a single cell. Conversely, imaging-based methods offer sub-cellular spatial resolution but is a targeted method that profiles hundreds to thousands of selected genes. Performing unbiased, genome-wide measurements and clearly delineating between cells to enable discoveries in complex tissues is much desired in spatial omics. We are developing a method that integrates the strengths of both imaging and sequencing to offer a novel approach to spatial omics with the potential for untargeted, genome-wide profiling of RNA, and other molecules, with single-cell resolution.

Our method utilizes light to encode spatial information onto nucleic acids and then use sequencing to determine genetic identity and spatial location of the molecules. This method tags RNA or proteins with oligos via reverse transcription or antibodies conjugated to oligos, respectively. Spatial information is encoded by assigning a unique barcode at each spatial location. A barcode is composed of a sequence of short oligonucleotides, which we refer to as letters. To stack letters to form a barcode, we use a light-directed ligation chemistry. Each letter has a photocleavable spacer that can be cleaved under UV illumination, allowing the next letter in the barcode to be recursively ligated. By spatially patterning the illumination, this process can photocleave multiple locations at once and is therefore multiplexable. Moreover, barcodes capable of error detection and correction can be used to improve spatial localization accuracy. Using this approach, we have demonstrated high efficiency of light-directed ligation of nucleic-acid letters, as well as spatial barcoding and sequencing.

Resolving dark genomic regions for clinical diagnostics using Oxford Nanopore long-read sequencing

Noemi Vidal-Folch[1], Guangchao Sun[2], Jagadheshwar Balan[3], Zachary D. Stephens[3], Ramanath Majumdar[1], Siva Arumugam Saravanaperumal[4], Qiliang Ding[1], Jesse R. Walsh[3], Lauren E. Schefter[3], Erik W. Klee[3], Jessica G. Drenzek[1], Makenzie L. Meyer[1], Ann M. Moyer[1], Linda Hasadsri[1], Aruna Rangan[1], Jennifer L. Herrick[1], Natasha T. Strande[1], Ross A. Rowsey[1], Jean-Pierre A. Kocher[3], Zhiyv Niu[1], Stephen J. Murphy[1], Devin Oglesbee[1]

1 Mayo Clinic, Department of Laboratory Medicine and Pathology, Rochester, MN
2 Sichuan Agricultural University, Maize Research Institute, Chengdu, Sichuan, China
3 Mayo Clinic, Department of Quantitative Health Sciences, Rochester, MN
4 Mayo Clinic, Genome Analysis Core, Rochester, MN

Introduction: Clinical diagnostics of genetic disorders is mainly performed by short-read NGS and supplemented with target-specific PCR-based methods for challenging medically relevant genes (CMRG). Supplemental testing increases turn-around-time, cost, and workflow complexity, while limitations remain on variant types, sizes, and haplotype phasing. These limitations hinder comprehensive characterization of CMRGs. In this study, we developed a long-read sequencing protocol to characterize single nucleotide variants (SNVs), insertions and deletions (indels), structural variants (SVs) and copy number variations (CNVs) in SMN1/SMN2, CYP21A2/CYP21A1P, RFC1 and DMD.

Methods: Human gDNA was extracted from 46 whole blood samples and 6 cell cultures using Puregene (Qiagen), Chemagic (Revvity) or Nanobind (Pacbio). gDNA was size-selected (SRE kit, PacBio), sheared to 30kb (Megaruptor3, Diagenode), and subjected to library preparation (SQK-LSK114, ONT). WGS was performed on PromethION R10.4.1 flow cells for 72h. Adaptive sampling with custom CMRG bed file was performed for SMN1/SMN2. Bioinformatics analysis included basecalling (Dorado), GRCh38 alignment (Winnowmap2), small variant calling (Clair3), SV calling (Sniffles2) and specialized in-house methods.

Improved isoform and fusion detection using full-length, whole-transcriptome sequencing on the MiSeq™ i100

Robin Bombardi[1], Richard Wang[1], Irina Khrebtukova[1], Shaveta Goyal[2], John Beckford[2], Andrew Farmer[2], Cande Rogert[1], Mark Wang[1], Rami Mehio[1], Emily Parker[1], Ali Crawford[1]

1 Illumina, Research & Development, San Diego, CA
2 Takara Bio USA, Inc., Research & Development, Mountain View, CA

SMART-Seq is a sequencing technique used to generate full-length, whole transcriptome cDNAs directly from cells or small amounts of bulk RNA. This approach uses SMART (Switching Mechanism at 5' End of RNA Template) technology to incorporate molecular barcoding information and universal adapters into the ends of each unique starting mRNA molecule through reverse transcription and Template-Switching Oligo (TSO)(1). SMART-Seq is commonly used to analyze gene expression, isoform expression, and gene fusion in a diverse array of contexts such as decoding the tumor micro-environment and understanding gene expression and function following CRISPR-Cas9 perturbation (2,3). Here, we perform MiSeq i100 sequencing on a library pool comprised of SIRVs (Spike-in RNA variant control mixes), myelogenous leukemia derived K562 cell line RNA, and mouse brain RNA, all prepared using the SMART-Seq v4 kit (Takara Bio USA). We demonstrate improvements in isoform detection and fusion calling with progressively longer read lengths on the MiSeq i100 instrument, from 2x150, 2x300, up to 2x500 read lengths. Our data demonstrates that greatly improved read-lengths and base quality results in a substantial increase in the frequency of uniquely barcoded, full-length whole transcriptome data that will enable researchers to better characterize and understand gene function.

Cell segmentation for complex neural cell morphologies

Lidan Wu[1], Aster Wardhani[1], Joseph-Tin Phan[1], Ashely Heck[1], Kimberly Young[1], Yi Cui[1], Stefan-Laural Rogers[1], Kathy Ton[1], Joseph Beechem[1]

1 Bruker Spatial Biology, Seattle, WA

Accurate cell segmentation is crucial for spatial omics, as incorrect cell boundaries can misassign transcripts or proteins, leading to skewed cellular profiles and flawed data interpretation. Brain tissue poses particular challenges due to its densely packed, irregularly shaped cells—many with star-like morphologies and long protrusions. Current machine-learning (ML) segmentation models often fail to capture these complex structures, as they are typically designed and trained on non-neuronal samples.

We present a novel cell segmentation pipeline designed for complex brain tissue morphology. The key innovations are a new neural-network model trained specifically for brain tissue and a custom post-processing method designed to recover large, irregular cells and their elongated structures, which are often missed by existing models.

Our segmentation model uses three color channels (nuclear, neuronal, and glial stains), rather than the standard two. This improves differentiation between closely positioned cells, such as astrocytes wrapped around neurons. The model was trained on a diverse dataset, including tissues and cultures, with balanced cell types to prevent bias towards dominant morphologies. Model outputs include spatial gradient flows and cell probability heatmaps. Initially, the default post-processing struggled to label large or elongated cells, like Purkinje neurons, and often missed long protrusions. To resolve this, we developed a custom post-processing method. After removing cells already recovered by default methods, we applied auto-thresholding and connectivity analysis to recover missed cells. For protrusions, we used distance-based cell-ID assignment to pixels in the model-predicted foreground.

With these advances, we successfully segment the complex morphologies of neural cells, including concave boundaries and long protrusions, where we find that expression profiles differ from those of cell bodies. In validation using CosMx™ data from FFPE human brain samples, we confirmed improved accuracy through both quantitative metrics and visual inspection. Our pipeline increased transcripts per cell by 17% with minimal change in cell volume, and reduced total cell number by 23% without obvious increase in extracellular transcript number, indicating improved boundary accuracy and fewer false positives.

Advances in this model and in our post-processing method offer insights that could extend to other challenging segmentation tasks, helping advance biological studies beyond brain tissue.

Utilizing SynBio to decipher the effect of promoter architecture on gene expression

Technology and Application Development

POSTER 471

Anna Yaschenko[1], Chengsong Zhao[1], Jose Alonso[1], Anna Stepanova[1]

1 North Carolina State University, Plant Biology, Raleigh, NC

Understanding gene expression regulation is central to the development of biotechnological solutions for several pressing agricultural problems. The advent of synthetic biology methods has opened the door to programmable genetic constructs that confer tunability and spatiotemporal regulation to gene expression. However, the relationship between promoter elements within the promoter of a gene and that gene's expression levels is not well defined. To explore this relationship, we have built synthetic promoters harboring up to ten transcription factor (TF) binding sites using GoldenBraid technology and subcloned these promoters upstream of a reporter gene. This was done by inserting one to ten copies of a protospacer, a 23bp recognition sequence for a dCas9-based synthetic TF, into a 1kb neutral promoter sequence that has no known TF binding sites in plants. We assembled the reporter gene to express a mCherry-Luciferase fusion protein as a readout and transiently co-expressed it along with a dCas9 activation system in Nicotiana benthamiana to test the effects of protospacer position, orientation, angular phase shift, copy number, and spacing on reporter expression. We find that protospacers in either the sense or antisense orientation confer comparable levels of gene expression and that the angular phase of the protospacer in relation to the transcription start site (TSS) has no prominent effect on reporter expression. Experiments testing protospacer position indicated that the ideal location of the TF binding site is about -150 base pairs (bp) upstream of the TSS and -119 upstream of the TATA box. Surprisingly, increasing protospacer copy number is linearly correlated with a boost in gene expression only for up to four copies of the protospacer, with a drop in reporter activity seen for five or more copies of this cis-element. Increasing spacing between protospacers was correlated with a boost in expression until about 123 bp, after which expression slowly decreased to the level of one protospacer suggesting that distant placement of the cis-element is not effective at supporting transcription initiation. Taken together, these results shed light on the rules of nature dictating promoter architecture that, in turn, ultimately pave the way to the creation of tunable expression systems.

Coalescing single-cell aneuploidy and transcriptomes at scale to decode breast cancer progression

Rui Ye[1,3,4^], Kaile Wang[1,2^], Shanshan Bai[1,2], Zhenna Xiao[1,2], Lei Yang[1,2,4], Chengling Tang[1], Emi Sei[1,2], Anna K. Casasent[1,2], Steven H. Lin[3], Alastair M. Thompson[5], Savitri Krishnamurthy[6], Nicholas E. Navin[1,2,7#]

1 University of Texas (UT) MD Anderson Cancer Center, Department of Systems Biology, Houston, TX
2 University of Texas (UT) MD Anderson Cancer Center, Department of Genetics, Houston, TX
3 University of Texas (UT) MD Anderson Cancer Center, Department of Radiation Oncology, Houston, TX
4 University of Texas (UT) MD Anderson Cancer Center, MD Anderson UT Health Graduate School of Biomedical Sciences, Houston, TX
5 Baylor University, Baylor College of Medicine, Dan L. Duncan Comprehensive Cancer Center, Department of Surgery, Houston, TX
6 University of Texas (UT) MD Anderson Cancer Center, Department of Anatomical Pathology, Houston, TX
7 University of Texas (UT) MD Anderson Cancer Center, Department of Bioinformatics, Houston, TX
^ These authors contributed equally.
# Correspondence author

Understanding the epithelial lineages of breast cancer and genotype-phenotype relationships requires direct measurements of the genome and transcriptome of the same single cells at scale. To achieve this, we developed wellDR-seq, the first high-genomic resolution, high-throughput method to simultaneously profile the genome and transcriptome of thousands of single cells in parallel. By benchmarking wellDR-seq, we found that both the DNA and RNA data quality show superior performance compared to other methods. We further profiled ~20,000 single cells from 6 estrogen receptor-positive breast cancer patients and identified ancestral cancer subclones in three patients that showed a luminal hormone responsive lineage, indicating a potential cell-of-origin. Our data also identified sporadic copy number aberrations (CNAs) in non-malignant cells including the two luminal epithelial lineages. Another unique aspect of wellDR-seq, is that it can directly quantify the impact of subclonal CNAs on gene expression programs in the cancer cells.

By quantifying the impact of subclonal CNAs on gene dosage, we found many (14.57-43.65%) expression differences occur outside of CNA regions in these tumors. Overall, these data reveal the complex relationships between chromosome aberrations and transcriptional programs in cancer cells, improving our understanding of breast cancer progression. Beyond cancer research, we anticipate that wellDR-seq will have broad applications in biomedical fields such as developmental biology, neuroscience, normal tissue mosaicism, aging and microbiology, where understanding the impact of genetic alterations on the phenotypes of a cell will advance our understanding of homeostatic biology and human diseases.

Natively integrating hybrid selection technology into the sequencing process and validating over a range of target panel sizes

Technology and Application Development

POSTER 473

Shawn Levy[1], Benjamin J. Krajacich[1], Kyle Metcalfe[1], Moa Hägglund[2], Carolina Diettrich[3], Mallet De Lima[3], Anna Lyander[2,3], Xiaodong Qi[1], Kelly Wiseman[1], Marina McCowin[1], Junhua Zhao[1]

1 Element Biosciences, San Diego, California, United States
2 Science for Life Laboratory, Department of Microbiology, Tumor and Cell Biology, Karolinska Institute, Stockholm, Sweden
3 Science for Life Laboratory, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, Stockholm, Sweden

Hybrid selection technologies allow highly efficient and accurate sequencing of genomic regions cost-effectively. As our understanding of the genome improved, demand has increased for tailored coverage of specific genes or regions using custom panels that span a variety of target sizes. Traditional methods involve probe hybridization followed by capture, washing, and amplification steps that must be performed prior to sequencing. Here we describe a novel target capture technology – Trinity™ that utilizes a one-hour hybridization followed by elimination of the standard post-hybridization steps and integration of those steps into the sequencing process. This technology integrates hybridization capture and sequencing. Based on the findings from exome panels, we examined the biochemical mechanisms underlying target capture dynamics and efficiency across panel sizes and characteristics. We also utilized the vast amount of available data to train machine learning algorithms to optimize and improve custom probe designs and predict output and performance. A matrix system was developed to support custom panels that demonstrated the capture response is correlated to the panel after non-linear transformation at logarithmic scale. We further validated 11 panels ranging from 7,000 to > 400,000 probes, from four major target capture panel vendors. Most of the panels tested using the predicted condition achieved ~ 87-94% on-Target and over 99% uniformity at the targeted coverage depth. Additional wet-lab optimizations further improved the coverage of extreme GC/AT regions in targets. Subsequently, we applied the new conditions to evaluate a 7,901-probe custom oncology panel. Using Trinity 1hr hybridization workflow, 10 samples including Horizon myeloid DNA reference standard and patient blood/bone marrow/fresh frozen tissue, were sequenced to achieve 3,884x-9,880x coverage. 22/22 cancer mutations in the standard were detected. For mutations with variant frequency 2.62% – 30%, observed allele frequency was concordant with expected allele frequency with R2 > 0.98. Notably, comparing Trinity 1hr hybridization and on-market workflow, we observed 1.76-fold lower duplications and corresponding 2.28-fold increase of library diversity being captured using Trinity technology, leading to potential further increase in assay sensitivity. Overall, the rapid workflow, broader compatibility with custom panels, and high sequencing quality of Trinity represents a highly novel technology for targeted sequencing.

Incorporation of unique molecular identifiers in PacBio Kinnex full-length RNA workflow

Huachun Zhong[1], Saranga Wijeratne[1], Jenell Lewis[1], Emilio Soriano Chavez[1], Maria E. Hernandez Gonzalez[1], James R. Fitch[1], Benjamin J Kelly[1], Elaine R. Mardis[1,2,3], Richard K. Wilson[1,2], Anthony R. Miller[1]

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

Long-read Isoform Sequencing (IsoSeq) protocols, such as those offered by Pacific Biosciences (PacBio), are valuable for exploring transcriptomic diversity. These workflows enable the sequencing of full-length transcripts without the need for read assembly, which provides a comprehensive view of alternative splicing, gene fusions, and novel isoform detection.

Previously, a limitation of IsoSeq was low read output per PacBio SMRT Cell, leading to potential misrepresentation of gene expression and missing low-abundance isoforms. However, with the introduction of the PacBio Revio system, coupled with the PacBio Kinnex workflow for concatenating full-length RNA transcripts, IsoSeq output has been significantly improved. The concatenation process, however, requires additional PCR amplification, raising concerns about PCR duplication, which can skew gene and isoform expression.

Unique Molecular Identifiers (UMIs) are used during data processing to remove duplicate transcripts that may arise from library preparation processes such as PCR amplification. To address the potential PCR duplication issue in the IsoSeq Kinnex workflow, we integrated a nine base-pair UMI into the template switching oligo (TSO) used during first-strand cDNA synthesis. This allows for the distinction between true biological molecules and amplification artifacts. We tested this by preparing two Kinnex full-length RNA libraries: one with the standard non-UMI TSO and the other with the UMI-enhanced TSO, using Universal Human Reference RNA as input. Both libraries were sequenced on separate Revio SMRT Cells for comparison.

Preliminary analysis utilized the PacBio read segmentation application, yielding 64,942,927 segmented reads (S-reads) from the non-UMI TSO sample and 61,545,234 S-reads from the UMI TSO sample. UMI deduplication revealed an approximate 41.7% duplication rate, with 35,847,246 s-reads remaining post-deduplication. Next, we utilized the SQANTI3 isoform classifier to evaluate isoform expression in both pre- and post-UMI deduplication samples. After filtering out transcripts categorized as IntraPriming, Mono-Exonic, RTSwitching, and LowCoverage/Non-Canonical, we focused on transcripts classified as full-splice match for our expression analysis. Significant expression changes were observed in 9,528 transcripts and 2,716 genes, with more than a two-fold difference when comparing the pre-UMI deduplication sample to the post-UMI deduplication sample. These results emphasize the crucial role of UMI incorporation in enhancing accuracy for differential gene and isoform expression studies.

A high-plexity exome solution tailored for sensitive variant detection in single cells

Technology and Application Development

POSTER 475

Durga M. Arvapalli[1], Laurel Coons[1], Tatiana V. Morozova[1], Jesse Ramirez[1], Subidhya Shrestha[1], Victor Weigman, Jon S. Zawistowski[1]

1 BioSkryb Genomics, Durham, NC

Whole exome sequencing (WES) is a mainstay and economical option to whole genome sequencing for identifying genomic variation but has classically been performed in a bulk context. Yet, single cell resolution is required for fully deconvolving subclonal evolution of mutation. To address this commercially underserved application we developed a single cell-specific exome solution, coupling complete and uniform genome amplification with single-cell optimized enrichment and sequencing parameters.

The foundation of this solution, ResolveXOME, is the enrichment of single-cell WGS libraries with a higher level of multiplexing (96) than standard exome enrichments that recommend 8-12 libraries per enrichment. We achieved the advantages of this higher plexity without compromising exome performance metrics. Critical for this performance was inputting faithfully-amplified single-cell genomes with ResolveDNA/OME, contrasting with existing studies that yield subpar exome target capture using classical WGA methodologies—due to allelic imbalance and partial genomic coverage.

We exemplified the performance of ResolveXOME with an enrichment plex mimicking a rare variant among a background of reference alleles. We spiked in a single leukemia cell with a drug resistance mutation in a background of single cells lacking this variant and showed the ability to discriminate this single nucleotide change despite the higher plexity of the enrichment. This was demonstrated both in the context of ResolveDNA genomic amplification and with ResolveOME multiomic amplification upstream of hybrid capture.

To carefully balance throughput and cost considerations with capture metrics, we arrived at a recommended sequencing targeting of 10M single (20M PE) reads per single-cell library in the enrichment. A mean target coverage of 25X was achieved, with both >80% of target bases covered at 10X and >80% on+near target coverage. Capture uniformity was demonstrated with a Fold80 penalty of 2.3. Critically, SNV calling sensitivity (>95%) and allelic balance (>90%), chief attributes of the Resolve amplification platforms, were maintained post-enrichment with ResolveXOME.

As a first single cell-centric exome solution with high plexity and robust upstream genomic amplification methodology, ResolveXOME can empower studies seeking variant detection in tumor samples, and when coupled with ResolveOME multiomics, the variant signature can be interpreted concurrently with definition of cell identity and phenotypic state.

Integrating long-read RNAseq and Next-Generation Protein SequencingÔ to explore proteoform variability

Gloria M. Sheynkman[1],[2],[3], Natchanon Sittipongpittaya[1], Kenneth A. Skinner[4], Erin D. Jeffery, Emily Watts-Whitehead[1]

1 University of Virginia, Department of Molecular Physiology and Biological Physics, Charlottesville, VA
2 University of Virginia, Department of Biochemistry and Molecular Genetics, Charlottesville, VA
3 University of Virginia, UVA Comprehensive Cancer Center, Charlottesville, VA
4 Quantum-Si, Branford, CT

A primary aim of proteogenomics is to couple the detection of distinct RNA isoforms with the expression of their cognate proteoforms. For example, the actin-binding tropomyosin (TPM1-4) genes produce over 40 spliceoforms, many of which are linked to complex diseases. However, subtle amino acid variations can complicate the detection of TPM proteoforms using standard proteomics methods. In this study, we present the first integration of long-read RNA sequencing (PacBio) to capture transcript isoforms, combined with PlatinumÒ (Quantum-Si), an innovative benchtop Next-Generation Protein Sequencing (NGPS) platform that produces single-molecule peptide sequencing data.

For this proof-of-principle study, we focused on TPM1 and TPM2, which share 87% amino acid identity. We derived a list of TPM1/2 isoform-informative peptides from our long-read transcript database. We then used Platinum to sequence synthetic peptides that correspond to TPM1/2 paralogs, spliceoforms, and post-translational modifications (PTMs). In the Platinum workflow, derivatized peptides are loaded into nanoscale reaction chambers of a semiconductor chip. During sequencing, dye-labeled N-terminal amino acid (NAA) recognizers reversibly bind to cognate NAAs, generating distinct binding and pulse duration (PD) patterns.

For the TPM peptides, we observed that Platinum can distinguish mixtures of peptides from tissue-specific TPM2 spliceoforms. We detected a 50% difference in PD between isobaric TPM1/2 peptides differing at leucine/isoleucine, which can be challenging to distinguish using mass spectrometry. We also observed a 9-fold difference in alignments between an unmodified and phosphotyrosine-containing peptide, indicating Platinum can detect PTMs. Taken together, our data establish the feasibility of applying NGPS to detect specific proteoforms at amino acid resolution, capturing all major forms of regulation—genetic, splicing, and post-translational—contributing to proteomic molecular diversity. We are now employing this platform to target proteoforms with genetic associations in complex human disease.

Validation of a novel autism spectrum disorder assay using whole-genome sequencing

Erin Connolly-Strong[1], Shruthi Sriam[1], Stephanie Ferguson[1], Johnny Chau[1], Alexis Tellez[1], June-Young Koh[1], Islam Oquz Tuncay[1], Sangmoon Lee[1]

1 Inocras Inc., San Diego, CA

Background:
Autism Spectrum Disorder (ASD) presents significant challenges in diagnosis and treatment due to its complex genetic underpinnings. The ASD Assay, developed by Inocras Inc., aims to detect pathogenic and likely pathogenic variants associated with ASD. The assay provides clinicians with valuable insights for early intervention and personalized treatment strategies. Additionally, the assay includes a polygenic score (PGS) for assessing the risk of ASD other neurological disorders. This study documents the analytical and clinical validation of this novel assay.

Methods:
The ASD Assay was validated using well-characterized cell lines and patient samples. The assay was performed on genomic DNA extracted from buccal swabs, saliva, buffy coat, or peripheral blood samples. The DNA was processed using the Watchmaker DNA Library Preparation Kit, sequenced on the Illumina NovaSeq X+ platform, and analyzed through the proprietary Inocras ASD pipeline. Analytical validation was conducted using six cell lines with known variants, including single nucleotide variants (SNVs) and insertions/deletions (indels). Clinical validation involved a cohort of previously diagnosed ASD patients (n=25), as well as a control group of negative germline samples. Variants detected by the assay were compared to the known clinical outcomes.

Results:
The ASD Assay demonstrated an overall sensitivity of 99.64% for SNVs and 96.46% for indels, with a positive predictive value (PPV) of 97.84% for SNVs and 92.60% for indels. In the clinical cohort, the assay successfully stratified positive and negative samples, confirming its ability to differentiate ASD-related variants.

Conclusion:
The ASD Assay demonstrates strong analytical performance in detecting variants associated with ASD, with high sensitivity and PPV. The assay's ability to stratify patients based on genetic variants and provide actionable insights for personalized treatment highlights its potential as a valuable tool in clinical settings for ASD management.

Olink®

Proteomics, Unplugged.

Powerful NGS-based proteomics— effortlessly uncover meaningful insights.

Olink Reveal makes multiplex proteomics accessible to every lab. Designed to seamlessly connect genomics and proteomics, Olink Reveal integrates with NGS, eliminating the need for new instrumentation or specialized expertise. Proteomics, unplugged from complexity - just accessible, actionable data.

Visit our suite in Osprey 8 today or www.olink.com/reveal to learn more

iconPCR

Presents...

iconPCR from n6:

The Next Frontier in Next Generation Sequencing

Scan for more details:

Join us @ Osprey 7
Calusa Ballroom

Attend our iconPCR Workshop

Tuesday, Feb 15 3:50pm - 4:05pm EST

Limits are for other people.

Scale single cell analysis up to **100 million** cells per run.

Stop by Caxambas-1 for the SEQ-EASY password.

PARSE BIOSCIENCES

Index

Index (continued)

Index (continued)

Index (continued)

Index (continued)

Index (continued)

Index (continued)

AGBT2025™

PRECISION HEALTH

SEPTEMBER 8 – 10, 2025

LOEWS CORONADO BAY | SAN DIEGO, CA

CALL FOR ABSTRACTS & REGISTRATION OPENS MAY 13, 2025

LEARN MORE AT AGBT.ORG

SCALE biosciences

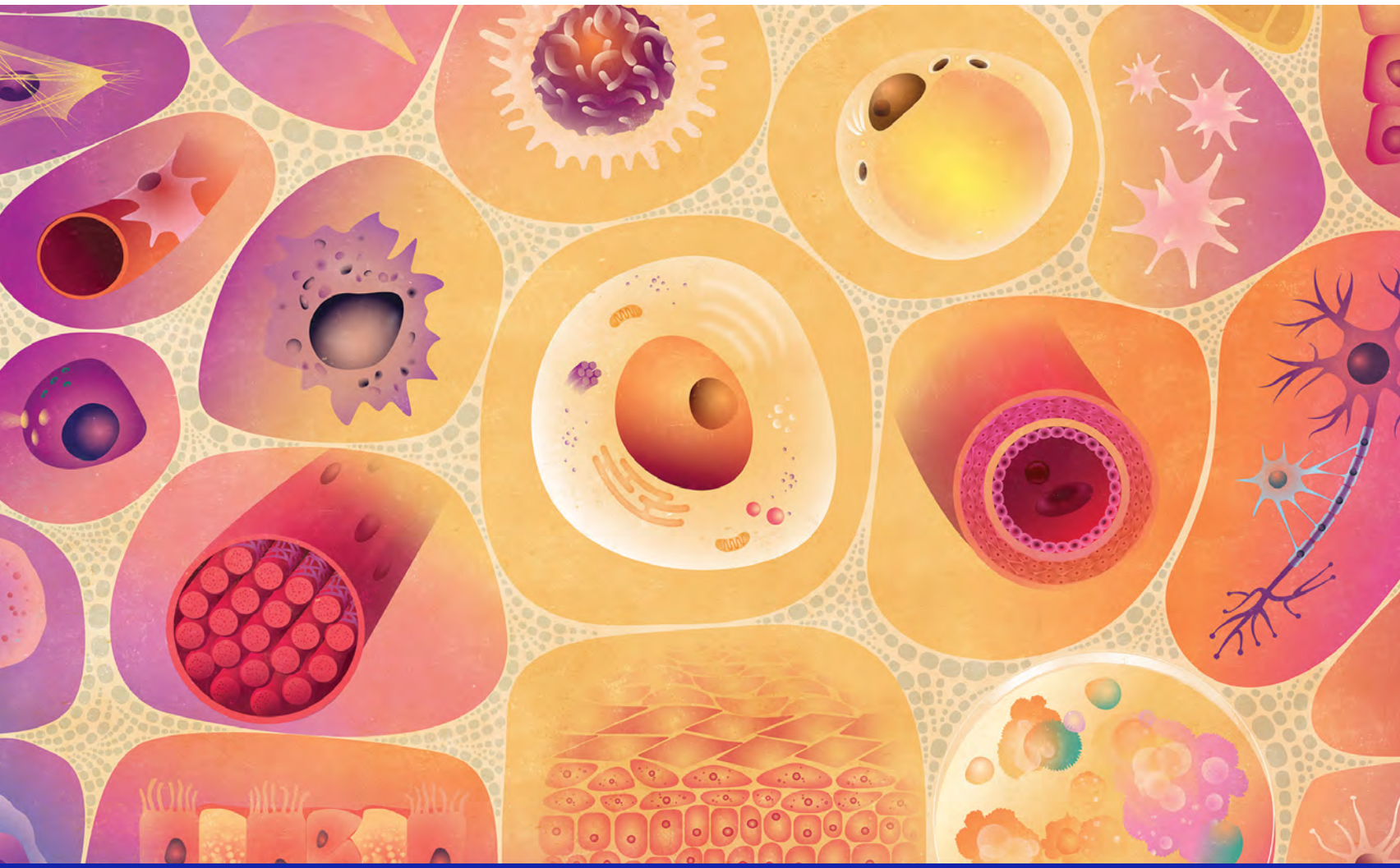
Turning Pieces Into Progress

Scale Bio was founded on a simple idea: empower scientists to pursue their most ambitious ideas without being held back by technical constraints. Now, QuantumScale is about to change everything you thought was possible in single cell genomics.

Learn how we're turning pieces into progress and revolutionizing the future of single cell at any scale.

Join our Bronze 8 talk:
Giovanna Prout, CEO
"Defy single cell gravity: Escape platform constraints in single cell"
4:20–4:32PM,
Calusa Ballroom 1–7

Join us in the Scale Bio Suite:
Every breakthrough starts with a single piece.
Caxambas 2



Find the missing piece in your single cell journey.

Stay Connected. Keep Innovating.

AGBT isn't just an event—it's a community.

The connections you make here don't end when the meeting does. Join us year-round to continue the conversations, collaborations, and breakthroughs that drive genomics forward.

Expand Your Impact

Stay at the forefront of discovery by attending AGBT Agriculture and AGBT Precision Health to connect across disciplines.

Deepen Your Network

The best ideas don't happen in isolation. Stay engaged with the scientists, innovators, and leaders shaping the future of genomics.

Exclusive Opportunities

Attending more than one AGBT meeting gives you access to unique networking, special promotions, and leadership opportunities in our community.

Don't be a "one and done" attendee—be part of something bigger!

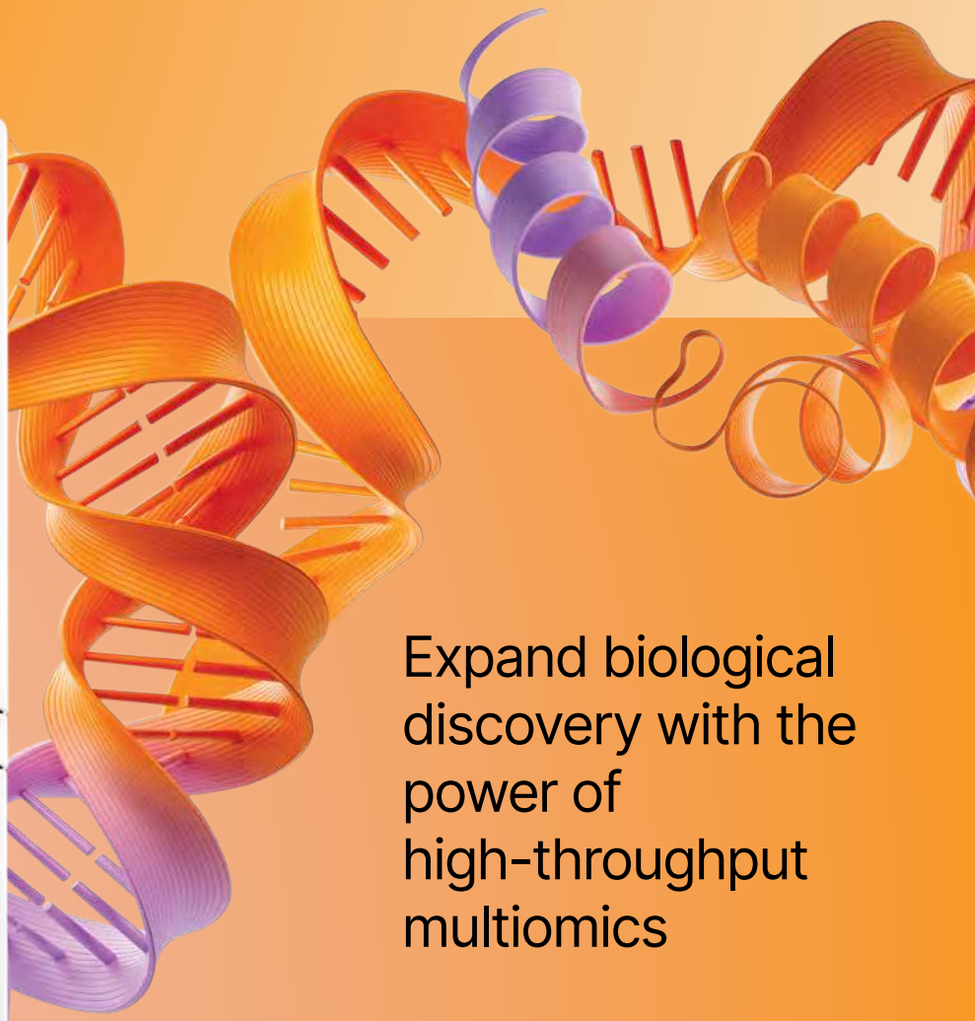
Learn more and stay connected.
Scan the QR code or stop by the registration desk.

PacBio

Awaken
to possibility

Visit us to see how
pacb.com/sequencing-systems

NovaSeq™ X Series: optimized for comprehensive omic solutions



Expand biological discovery with the power of high-throughput multiomics

To learn more, scan the code to the left or visit:
https://ilmn.ly/novaseq-x-multiomics

illumina® | This is the Genome Era™