Canadian Bioinformatics Hub Conference

2026

# Welcome to CBHC!

Canadian Bioinformatics Hub Conference

2026

| Wednesday May 27: Workshops | | |
|---|---|---|
| **Start Time** | **End Time** | **MaRS Collaboration Center** |
| **Sponsored Workshop - Unlocking Cellular Insights: AWS Cloud Solutions for Advanced Omics and Generative AI Research** | | |
| 08:00 | 08:45 | Breakfast and Registration |
| 09:00 | 09:15 | Welcome & Introduction to AWS Open Data |
| 09:15 | 10:00 | Module 1: Accessing Data from AWS Open Data |
| 10:00 | 10:45 | Module 2: Lightsail for Research |
| 10:45 | 11:30 | Module 3: Genomics Use Case with Kiro |
| 11:30 | 12:00 | Other use cases and resources |
| *12:00* | *13:00* | *Lunch & Networking Session with all workshops* |
| | | |
| **Workshop #1 - Bioinformatics as a Career: An insider view** | | |
| 08:00 | 08:45 | Breakfast and Registration |
| 09:00 | 09:15 | Welcome & Introductions |
| 09:15 | 10:15 | Interactive Panel Discussion with Trainee Presenters |
| *10:15* | *10:45* | *Coffee Break & Networking* |
| 10:45 | 12:15 | Hands-on activity: Designing your career path |
| *12:15* | *13:00* | *Lunch & Networking Session with all workshops* |
| | | |
| **Workshop #2 - Nurturing an Early Career growth ecosystem in Canada** | | |
| *12:00* | *13:00* | *Registration + Lunch & Networking Session with all workshops* |
| 13:00 | 13:15 | Welcome & Introductions |
| 13:15 | 14:30 | Setting the Stage - Early Career Professional Perspectives |
| *14:30* | *15:00* | *Coffee Break & Networking* |
| 15:00 | 16:15 | Breakout Groups - Guided Discussions on Resourcing BCBDS Careers |
| 16:15 | 17:00 | Report Back & Next Steps |
| | | |
| **Workshop #3 - Reproducible Research: Essentials for Managing Your Data** | | |
| *12:00* | *13:00* | *Registration + Lunch & Networking Session with all workshops* |
| 13:00 | 13:15 | Welcome & Introductions |
| 13:15 | 13:30 | Module 1: Introduction to Research Data Management (RDM) |
| 13:30 | 14:30 | Module 2: Version Control with Git |
| *14:30* | *15:00* | *Coffee Break & Networking* |
| 15:00 | 16:00 | Module 3: Organizing Your Data for Machine-Actionability |
| 16:00 | 17:00 | Module 4: Facilitating Discovery of Your Datasets using Open Source Platforms and Standards |

| Thursday, May 28 (Day 1) | | | |
|---|---|---|---|
| **Start Time** | **End Time** | **MaRS Auditorium/Concourse** | |
| 08:00 | 09:00 | Registration | |
| 09:00 | 09:15 | Land Acknowledgement & Welcome Address | Dr. Michelle Brazas, Bioinformatics.ca |
| 09:15 | 09:25 | Opening Remarks - Positioning Bioinformatics in Canadian Ecosystem | Naveed Aziz, Genome Canada and Chris McMaster, CIHR |
| 09:30 | 10:30 | Keynote | Dr. Robert Gentleman, Dana Farber Cancer Institute |
| *10:30* | *11:00* | *AM Coffee Break* | |
| **Session 1 : Data initiatives and resources**<br>Chair: Drs. Emma Griffiths, Finlay Maguire, and Guillaume Bourque | | | |
| 11:00 | 11:30 | Plenary talk | Dr. Mélanie Courtot, Ontario Institute for Cancer Research |
| 11:30 | 11:45 | Short talk | Selected from abstracts |
| 11:45 | 12:00 | Short talk | Selected from abstracts |
| 12:00 | 12:15 | Flash talks (3min/speaker; 5 speakers) | Selected from abstracts |
| 12:15 | 13:00 | Panel Discussion: Scaling up with National Computing Infrastructure | Dr. Guillaume Bourque, Digital Research Alliance of Canada<br>Dr. Pablo Prieto Barja, Lifebit<br>Dr. Fiona Brinkman, Simon Fraser University, Dr. Mélanie Courtot, OICR |
| *13:00* | *14:30* | *Lunch Break & Poster Session #1 - Even Posters (Jewelbox Room, main floor)* | |
| **Session 2: Computational methods and algorithms**<br>Chairs: Dr. Arvind Mer | | | |
| 14:30 | 15:00 | Plenary talk | Dr. Anamaria Crisan, University of Waterloo |
| 15:00 | 15:15 | Short talk | Selected from abstracts |
| 15:15 | 15:30 | Short talk | Selected from abstracts |
| 15:30 | 15:45 | Flash talks (3min/speaker; 5 speakers) | Selected from abstracts |
| *15:45* | *16:00* | *PM Coffee Break* | |
| 16:00 | 16:45 | Fireside Chat: Do we really need another bioinformatics tool? To build or not to build | Hamed Najafabadi, McGill University<br>Benjamin Haibe-Kains, UHN |
| 16:45 | 17:00 | Short talk | Sidki Bouslama, Illumina |
| 17:00 | 17:45 | Success Circles - Thought Leadership Conversations | |
| 18:00 | 20:00 | Networking Mixer | |

| Friday, May 29 (Day 2) | | | |
|---|---|---|---|
| **Start Time** | **End Time** | **MaRS Auditorium/Concourse** | |
| 08:00 | 09:00 | Registration | |
| **Session 3: Leveraging data and methods to discover biological insights**<br>Chairs: Drs. Mohamed Helmy, Finlay Maguire, and Guillaume Bourque | | | |
| 09:00 | 09:30 | Quantum computing applications in biology | Dr. Steven Rayan, QuanTA, U. Saskatchewan |
| 09:30 | 09:45 | Short talk | Selected from abstracts |
| 09:45 | 10:00 | Flash talks (3min/speaker; 5 speakers) | Selected from abstracts |
| 10:00 | 10:45 | Panel Discussion: Building Sustainable Funding for Bioinformatics | Dr. Josette-Renée Landry, Génome Québec<br>Dr. Fei-Fei Liu, University of Toronto / CIHR Institute of Cancer Research<br>Dr. Michael Sefton, University of Toronto / MedByDesign<br>Dr. Andrew McArthur, McMaster University |
| *10:45* | *11:15* | *AM Coffee Break* | |
| 11:15 | 11:45 | Computational Genomics | Dr. Claudia Kleinman, SofM, McGill University |
| 11:45 | 12:00 | Short talk | Selected from abstracts |
| 12:00 | 12:15 | Flash talks (3min/speaker; 5 speakers) | Selected from abstracts |
| 12:15 | 13:00 | Fireside Chat: Human vs machine in research discoveries | Jérôme Waldispühl, McGill University<br>Rachel Edgar, University of Toronto |
| *13:00* | *14:30* | *Lunch Break & Poster Session #2 - Odd Posters (Jewelbox Room, main floor)* | |
| **Session 4: Recognizing excellence in Canadian bioinformatics**<br>Co-Chairs: Drs. Gary Bader, Aline Talhouk, Arnaud Droit | | | |
| 14:30 | 14:35 | Award introduction & overview | Gary Bader, Aline Talhouk, Arnaud Droit, Francis Ouellette |
| 14:35 | 15:00 | CBH Award distribution | Recipients for CBH Research & Innovation and Francis Ouellette Community Awards |
| 15:00 | 16:00 | Keynote | Dr. Amy Lee, Simon Fraser University |
| 16:00 | 16:10 | Let's talk about IDEAS | Dr. Julie Hussin, Université de Montréal |
| 16:10 | 16:30 | Plenary talk: Looking ahead at CBH | Dr. William Hsiao, SFU |

# Canada's National Federated Research Infrastructure.

CanPath

lifebit

Lifebit and CanPath have established a national framework that lets researchers collaborate across institutions while data stays secure and under local control — **hosted in Canadian cloud environments.**

> "As more partners join, the benefits grow for researchers, institutions, and ultimately for Canadians."
>
> Dr. Philip Awadalla · National Scientific Co-Director, CanPath

**350,000+**
Participant volunteers across Canada

**1M**
Data points and growing

**178,000+**
Biosamples of DNA source material

**Learn more**
canpath.ca
lifebit.ai

# How AWS supports researchers

**Share results – Community building**

- AWS Summits
- Case studies, blogposts
- Workshops
- GitHub – Share code

**Accelerate time to science**

- Research specific solutions
- Data egress waiver
- Share data – AWS Open Data (Reach out to opendata@amazon.com)

aws open data

**Research proposals**

- Research proposal support
- Letters of support
- AWS related document support

**Build research on AWS**

- Simplified access / Secure infrastructure
- Get trained on AWS / Immersion days
- Partner solutions
- AWS cloud credits for research

## Thursday, May 28th

**09:00-09:15**

**Land Acknowledgement & Welcome Address**

**09:15-09:25**

**Opening Remarks by Genome Canada/CIHR**

**Presentation Overview:**
Positioning Bioinformatics in Canadian Ecosystem

**09:25-10:30**

**Keynote - Opportunities in High Dimensions**

**Authors List:**
**Robert Gentleman**, Dana Farber Cancer Institute, USA

**Presentation Overview:**
I will discuss a number of different approaches and methods to help explore high dimensional molecular data. A primary tool will be principal component analysis (PCA). I will show how it can be used to explore data integration, cell type assignment and explore a variety of different assessments of spatial transcriptomic and proteomic data. I will use some examples from the scDiagnostics package (https://github.com/ccb-hms/scDiagnostics) and its companion paper (https://www.biorxiv.org/content/10.64898/2026.01.29.701618v1).

**10:30-11:00**

**Morning Coffee Break**

**11:00-11:30**

## Does AI dream of structured datasets?

**Authors List:**
**Mélanie Courtot**, Ontario Institute for Cancer Research & University of Toronto, Canada

**Presentation Overview:**
Realizing the promise of AI in genomics depends on a simple assumption: the availability of clean, structured, interoperable data. Yet achieving this across diverse datasets, systems, and organizations remains one of the field's most persistent challenges. In this talk I argue that neither technology nor standards alone are sufficient and that to achieve real-world impact we must co-design tools, data standards, and governance frameworks to work in concert. Drawing on two complementary case studies, I will explore what this looks like in practice. In human genomics, the ICGC ARGO program demonstrates how large-scale, multi-institutional cancer genomics initiatives can achieve meaningful data harmonization while respecting the ethical and consent obligations owed to participants worldwide. In microbial genomics, iMicroSeq illustrates the distinct but equally complex interoperability challenges that arise in pathogen surveillance and environmental sequencing contexts. Together, these examples reveal shared patterns: the need for shared vocabularies and ontologies, scalable data quality pipelines, federated access models, and stakeholder alignment that spans clinical, research, and public health communities. The talk closes by reflecting on where AI genuinely helps and where it exposes the limits of poorly structured data, making the case that investing in data infrastructure is a prerequisite for AI to deliver on its potential.

**11:30-11:45**

## Long-read proteogenomic atlas of human neuronal differentiation reveals isoform diversity informing neurodevelopmental risk mechanisms

**Authors List:**
Melanie Davie, University of Toronto, Canada
Brett Trost, University of Toronto, Canada
Shreejoy Tripathy, University of Toronto, Canada
Craig Smibert, University of Toronto, Canada
Howard Lipshitz, University of Toronto, Canada
Hyun Lee, University of Toronto, Canada
Julien Muffat, University of Toronto, Canada
John Calarco, University of Toronto, Canada
Yun Li, The Hospital for Sick Children, Canada
Fatima Naimi, The Hospital for Sick Children, Canada
**Nuo Xu**, University of Toronto, Canada
Maria Eleni Fafouti, University of Toronto, Canada
Ellie Hogan, University of Toronto, Canada
Hua Luo, University of Toronto, Canada
Jimmy Ly, Whitehead Institute for Biomedical Research, United States
Chaoying Long, The Hospital for Sick Children, Canada
Maahil Arshad, University of Toronto, Canada
Ai Tian, University of Toronto, Canada
Elizabeth Radley, University of Toronto, Canada
Katherine Rynard, University of Toronto, Canada

**Presentation Overview:**
RNA splicing shapes neuronal identity and disease risk, yet current maps lack the developmental resolution and depth to resolve this complexity. Here, we integrate deep long-read RNA sequencing and proteomics in iPSC-derived cortical neurons to generate a high-resolution proteogenomic atlas of human neuron development. We identify 182,371 mRNA isoforms (over half previously unknown) and provide direct peptide evidence for the translation of hundreds of novel protein-coding sequences. Population genetics demonstrates that variants affecting novel exons and splice sites are under negative selection, underscoring the potential significance of these isoforms. During neuronal maturation, we observe that ASD risk genes undergo dynamic isoform switching, including microexon inclusion and intron retention, that remodel key protein domains and regulatory regions. Furthermore, we uncover widespread, long-range coordination between splicing and polyadenylation. Finally, our atlas enables variant reinterpretation in ASD, highlighting the value of an isoform-centric view for interpreting pathogenic variation in neurodevelopment.

**11:45-12:00**

**Integration of diverse multiomics, environmental and clinical data from the CHILD Cohort Study, into the CHILDdb platform, enables more wholistic insight into childhood asthma etiology**

**Authors List:**
Lauren Erdman, The Hospital for Sick Children, Canada
Fiona Brinkman, Simon Fraser University, Canada
Padmaja Subbarao, The Hospital for Sick Children, Canada
Stuart Turvey, University of British Columbia, Canada
Meghan Azad, University of Manitoba, Canada
Theo Moraes, The Hospital for Sick Children, Canada
Jeffrey Brook, University of Toronto, Canada
Catherine Field, University of Alberta, Canada
Kozeta Miliku, University of Toronto, Canada
Michael Kobor, University of British Columbia, Canada
**Geoff Winsor**, Simon Fraser University, Canada
Stan He, The Hospital for Sick Children, Canada
Maxwell Libbrecht, Simon Fraser University, Canada
Karar Al-Mamaar, Simon Fraser University, Canada
Luisa Mercada Mendoza, Simon Fraser University, Canada
Charisse Petersen, University of British Columbia, Canada
Karlie Edwards, University of British Columbia, Canada
Erin Gill, Simon Fraser University, Canada
Justin Cook, Simon Fraser University, Canada

**Presentation Overview:**
The CHILD Cohort Study is a longitudinal population-based birth cohort, following 3,454 healthy Canadian children and their families from early pregnancy. The open-source CHILDdb.ca platform integrates over 24,000 variables, plus multiomics datasets, with an interface enabling secure, privacy-preserving discovery and visualization of complex data, while streamlining access requests. Integrated data includes questionnaires and clinical assessments, household- and neighbourhood-level exposures, and sample-derived chemical analyte datasets. An exposome-wide association study using 2,954 early-life exposures was used to identify exposures associated with childhood asthma, the most common chronic disease among children. Asthma-associated exposures included prenatal cleaning product exposure, breast milk metabolites, phthalates, antibiotic use, PM2.5 and ozone – with machine learning providing evidence for interaction between air quality measures and inherited factors. Further mechanistic investigation revealed downstream changes dependent on exposure, including epigenetic changes in cord blood, and dysbiosis of children’s gut microbiomes. The results suggest a complex etiology for asthma and highlight the value of employing unbiased approaches that utilize both prenatal and postnatal environmental factors, together with multiomics data. CHILDdb, through consistent, harmonized data integration from multiple Canadian sites, can be applied toward better understanding child health trajectories and investigating responses to disease - informing clinical practice and public health policy.

**12:00-12:15**

## Preparing data for a long journey: early curation helps researchers generate more reusable, high quality datasets

**Authors List:**
**Ruobin Liu**, University of Saskatchewan, Canada
Lacey-Anne Sanderson, University of Saskatchewan, Canada
Carolyn T. Caron, University of Saskatchewan, Canada
Reynold Tan, University of Saskatchewan, Canada
Vidusha Wilpita, University of Saskatchewan, Canada
Laura Jardine, University of Saskatchewan, Canada
Kirstin Bett, University of Saskatchewan, Canada

**Presentation Overview:**
Data is a fundamental element in research of any kind. While large data repositories exist for genomics data, how can researchers better support all stages of a research experiment, from generation, backup, publication, and reuse? Some challenges we have encountered include inconsistencies in data collection method, ambiguity in links between datasets, and miscommunication on custom file formats. Our work takes a wide range of data inputs including, but not limited to, raw agricultural field trials, biochemical assays, and high throughput imaging outputs. These data are backed up during active collection and then curated both automatically via validation tools built into database importers, and manually by a curator in collaboration with the data collectors. This ensures consistency and completeness of the data, including metadata, prior to being stored in a relational database with FAIR principles in mind. Furthermore, having these high-quality datasets stored in a database allows seamless integration with bioinformatic and data visualization tools, generation of dynamic pages describing datasets and facilitating deeper insights into research data through future reuse. The resulting system not only enhances current biological insight but also creates a sustainable foundation for future research, collaboration, and innovation.

**Chicken or Egg: did immune ligands evolve before their receptors?**

**Authors List:**
**Colin O'Dwyer**, Ottawa Hospital Research Institute, Canada
David Cook, Ottawa Hospital Research Institute, Canada
Michele Ardolino, Ottawa Hospital Research Institute, Canada

**Presentation Overview:**

Due to its short cytoplasmic tail, PD-L1 has historically been considered merely an inert ligand for its canonical receptor PD-1. However, emerging evidence illustrates that PD-L1 possesses a range of cell-intrinsic functions independent of PD-1. Given these intrinsic roles in the absence of its receptor, we investigated whether these functions represent ancestral roles from before PD-L1 evolved as a ligand for PD-1. We developed an evolutionary analysis pipeline that aligns human proteins to animal proteins across 24 curated databases spanning Metazoa. The pipeline identifies the best aligning animal proteins and derives information about their conserved domains, synteny, and structural similarities to the human protein of interest. Using this approach, we deduced the evolutionary conservation of PD-L1 and PD-1. Our analysis suggests that PD-L1 appears earlier in evolutionary time than PD-1, potentially explaining why PD-L1 retains cell-intrinsic effects even in the absence of its receptor. Additionally, we found evidence suggesting that many immune ligands evolved before their cognate receptors, highlighting an intriguing broader evolutionary pattern. Positioned at the intersection of comparative immunology and evolutionary biology, this work addresses gaps in our fundamental understanding of immune proteins and provides evolutionary context for the multifunctional nature of immune checkpoint molecules.

**GenPipes: a workflow manager for accessible and reproducible bioinformatics**

**Authors List:**
**Mareike Janiak**, McGill University / Canadian Centre for Computational Genomics, Canada
Paul Stretenowich, McGill University / Canadian Centre for Computational Genomics, Canada
Robert Eveleigh, McGill University / Canadian Centre for Computational Genomics, Canada
Jean-Michel Garant, McGill University / Canadian Centre for Computational Genomics, Canada
Alexa Li Kam Wa, McGill University / Canadian Centre for Computational Genomics, Canada
Jose Hector Galvez Lopez, McGill University / Canadian Centre for Computational Genomics, Canada
Guillaume Bourque, McGill University / Canadian Centre for Computational Genomics, Canada

**Presentation Overview:**
GenPipes is an open-source Python-based workflow management system offering 10 pre-built bioinformatics pipelines. Its features include smart restarts, easy configuration, support for multiple job schedulers and cloud computing, detailed logs and documentation. An interactive command-line wizard provides guidance for beginners, while ticket support is available to all users. GenPipes is integrated with the clusters of the Digital Research Alliance, where it can be loaded as a module and accesses a central software stack, so users do not need to handle any installations. Pre-built pipelines are available for RNA-seq, long-read sequencing, and WGS, among others, and include protocols such as paired tumour/normal analysis. This protocol has been used to analyze more than 4400 cancer cases for the Marathon of Hope project in Québec, illustrating its scalability and applicability to large projects. Continuous improvements are made to pipelines in response to user feedback and internal benchmarking. Current development goals include better support for the T2T human reference, enhancing of actionable reporting for somatic variants, and improving pipelines for long-reads. GenPipes is of interest to a broad user base, as it both lowers the barrier of entry to bioinformatic analysis and ensures research reproducibility, crucial for large, multi-year projects.

# The Canadian Genomic Data Commons (CGDC): A National Infrastructure for Genomic Data Sharing

**Authors List:**
**Erika Frangione**, Mount Sinai Hospital, Sinai Health, Canada
Jordan Lerner-Ellis, Mount Sinai Hospital, Sinai Health, Canada
Navneet Aujla, Mount Sinai Hospital, Sinai Health, Canada
Heidi Rehm, Broad Institute, United States
Marc Fiume, DNAstack, Canada
Vincent Ferretti, Université de Montréal, Canada
Yann Joly, McGill University, Canada
Steven Jones, BC Cancer, Canada
Patrick Frosk, University of Manitoba, Canada
Sherryl Taylor, University of Alberta, Canada
Kym Boycott, University of Ottawa, Canada

**Presentation Overview:**
Clinical genome sequencing (GS) generates large data volumes with prohibitive storage costs, and remains largely inaccessible to researchers and clinicians. The Canadian Genomic Data Commons (CGDC) will use high-performance computing (HPC) to securely transfer genomic data from healthcare settings into a data sharing ecosystem. This is especially urgent for rare diseases (RDs), which afflict >3 million Canadians. The CGDC will develop three core facilities to create sustainable, federated, and secure digital infrastructure for data sharing across Canada: 1. The Canadian Open Genetics Repository (COGR); 2. Canadian genome aggregation database (gnomAD-Canada); and, 3. Tools for RD researchers. 1) COGR will centralize the standardization and sharing of variant data from diagnostic laboratories through laboratory-specific automated data transfer pipelines, variant discrepancy reporting, and submission of consensus interpretations to COGR and ClinVar. 2) A Canadian instance of the gnomAD browser will be established using aggregated allele frequencies from Canadian large-scale GS projects. 3) A Canadian deployment of the seqr platform on HPC4Health will support gene–disease discovery, alongside a dynamic consent portal developed in collaboration with the Pan-Canadian Genome Library (PCGL). The CGDC will expand HPC research capacity, improve variant interpretation and quality assurance, and maximize the value of Canadian clinical and research cohorts.

**Sustainability, Intellectual Property, and theComprehensive Antibiotic Resistance Database**

**Authors List:**
**Andrew McArthur**, McMaster University, Canada
Brian Alcock, McMaster University, Canada
Amogelang Raphenya, McMaster University, Canada

**Presentation Overview:**
Increasing rates of antimicrobial resistance (AMR), i.e., drug-resistant infections, is a direct threat to modern medicine. In Canada, approximately 26% of infections in 2018 were resistant to the drugs generally used to treat them. Globally, current rates of AMR and predictions for increased rates of DRI are equally dire. Yet, antimicrobial resistance in human populations does not occur in isolation but instead occurs within the broader context of antimicrobial use in agriculture and the pollution of natural environments. Canada has instituted a Federal Action Plan on Antimicrobial Resistance and Use, of which one focus is surveillance of AMR. The Comprehensive Antibiotic Resistance Database (card.mcmaster.ca) is a made-in-Canada database and software platform used around the world for genomic surveillance of antimicrobial resistance genes (ARGs) by applying diverse data science approaches from genomics to natural language processing to protein language models. Although an academic enterprise, CARD by design does not rely on grant funding for sustainability and support of professional biocurators and bioinformaticians. CARD's data and resources are free for academics and public health agencies but require a license for private sector use, requiring a careful balance of academic research, service to the broader community, dedication to open source, and commercialization.

**12:15-13:00**

**Scaling up with National Computing Infrastructure**

**14:30-15:00**

## Bridging the Last Mile: Why Human-Centered Design Matters in Bioinformatics

**Presentation Overview:**

Bioinformatics has a last mile problem: we build increasingly sophisticated methods that rarely translate into real-world clinical or research use. The issue is not a lack of sophisticated algorithms. It is a lack of attention to the humans expected to use them. Tools are often designed around data and performance metrics, rather than the workflows, constraints, and decision-making contexts of their end-users. In this talk, I explore how human-centered design principles can be applied to bioinformatics. Through two case studies, I demonstrate how this perspective reshapes data discovery bottleneck. The first presents automated approaches for identifying dataset linkages and automatically analyzing and visualizing connected data. The second expands this idea further to demonstrate how a multi-agent AI approach can be harnessed to integrate diverse bioinformatics tools into both research and clinical use. In each case, I show how incorporating user perspectives not only improved accessibility and interpretability, but also changed what questions were asked and how results were ultimately applied. Together, these examples show that human-centered design is not an add-on, but a driver of how bioinformatics tools can be conceived, built, and used. I conclude by discussing how the last mile problem will require more than better algorithms. It will require rethinking how we design tools in the first place: around the people and contexts in which they are actually used.

**15:00-15:15**

## Clonal-Resolution Modeling of hPSC Population Dynamics for Quality Control in Cell Therapy

**Authors List:**
**Nava Leibovich**, National Research Council Canada, Canada
Ali Shahdoost, The School of Biomedical Engineering, The University of British Columbia, Vancouver, Canada, Canada
Nika Shakiba, The School of Biomedical Engineering, The University of British Columbia, Vancouver, Canada, Canada

**Presentation Overview:**
Human pluripotent stem cell (hPSC)-derived therapies require safe, large-scale expansion. However, the emergence of genetic variants during manufacturing poses significant risks to product safety and efficacy. Understanding variant behavior is critical for robust quality control. We developed an empirically informed in-silico model of hPSC growth that captures stochastic interplay between wild-type and variant populations. The framework incorporates cell sampling, variant growth advantages, and passaging effects to simulate diverse manufacturing scenarios, including scale-up (increasing bioreactor volume) versus scale-out (increasing the number of bioreactors). Simulations reveal critical trade-offs between contamination risk and production costs. Notably, our clonal-resolution modeling generates phenotypic population signatures that function as a novel tool for detecting contamination events. Preliminary data suggest that this method enables the early detection of variant expansion, allowing for timely intervention in the production pipeline. This work advances computational methods for cell therapy manufacturing by providing a data-driven resource for pipeline optimization. By integrating quantitative modeling with biological insights, we offer a scalable approach to mitigate variant contamination, ensuring the safe delivery of hPSC-based products. This tool establishes a foundation for automated quality control in regenerative medicine workflows

**15:15-15:30**

## An Integrated Pipeline for Enhanced Viral Discovery in Wastewater Metatranscriptomes Through Targeted Depletion of Tobamoviruses and Ribosomal RNAs

**Authors List:**
Maria A. Bautista, Department of Biological Sciences, University of Calgary, Canada
**Tulika Bhardwaj**, Department of Biological Sciences, University of Calgary, Canada
Dylan J. Barbera, QIAGEN Sciences Inc, Maryland, United States
Samuel J. Rulli, QIAGEN Sciences Inc, Maryland, United States
Janine McCalder, Department of Biological Sciences, University of Calgary, Canada
Chloe Papparis, Department of Biological Sciences, University of Calgary, Canada
Xiaoli Pang, Department of Laboratory Medicine & Pathology, Faculty of Medicine & Dentistry, University of Alberta, Canada
Bonita Lee, Department of Pediatrics, Faculty of Medicine & Dentistry, University of Alberta, Canada
Rhonda G. Clark, Department of Biological Sciences, University of Calgary, Canada
Nicole Acosta, Department of Microbiology, Immunology, and Infectious Diseases, University of Calgary, Canada
Christine O' Grady, Advancing Canadian Water Assets, Calgary, Canada
Michael D. Parkins, Department of Microbiology, Immunology, an

**Presentation Overview:**
Wastewater metatranscriptomic datasets are inherently compositionally biased, which diminishes the detection of low-abundance RNA viruses and complicates downstream taxonomic and genome-centric analyses. Sequencing reads are dominated by ribosomal RNAs (rRNAs) and highly abundant plant-infecting tobamovirus RNA genomes derived from human diets. To overcome these challenges, samples from a wastewater treatment plant in Calgary, Alberta were processed using a depletion strategy employing rRNA-targeting probes QIAseq FastSelect 5S/16S/23S together with custom designed Tobamovirus RNA probes. An integrated bioinformatics workflow was developed by leveraging complementary computational tools to maximize viral signal recovery. Residual rRNA reads not removed by the depletion probes were detected and removed using RiboDetector, followed by taxonomic classification of non-rRNA reads using Kraken2. Applying geNomad identified viral sequences, enabling sensitive detection of both known and divergent viral genomes. Genome quality, completeness, and contamination were assessed using CheckV, ensuring robust downstream interpretation. Benchmarking this approach against datasets generated without these depletion probes demonstrated a 90% increase in classified viral reads, with pronounced enrichment of Riboviria sequences. Despite persistent challenges in extracting low-abundance signals from complex mixtures, this study highlights the potential of targeted depletion to enhance the sensitivity and cost-effectiveness of wastewater-based surveillance for viral pathogen monitoring.

**15:30-15:45**

## Graph Attention Networks for Chromosome-Level Modeling of cfDNA Methylation in Cancer Detection

**Authors List:**
**Lin Zhao**, Algoma University, Canada
Vanessa Giuliano, Algoma University, Canada
Yan Yan, University of Guelph, Canada
Wenjun Lin, Algoma University, Canada
Ping Luo, Algoma University, Canada

**Presentation Overview:**

Recent advances in cell-free DNA (cfDNA) methylation profiling have enabled minimally invasive cancer detection using plasma samples, providing a promising alternative to conventional tissue-based diagnostics. Existing approaches typically combine cfDNA methylation signals with statistical feature selection and machine-learning classifiers to identify cancer samples. However, these methods often restrict learning to top-ranked differentially methylated regions (DMRs), potentially overlooking informative signals, and rely on fixed-size genomic bins that fail to capture DMRs with their true variable lengths. As a result, complex methylation patterns and long-range genomic dependencies at the chromosome scale are insufficiently modeled. To address these limitations, we propose a graph-structured modeling framework that explicitly incorporates genomic locations and relationships among chromosomal bins, enabling a biologically informed representation of cfDNA methylation data. Chromosomes are modeled as a graph, with nodes representing genomic bins and edges encoding positional and contextual relationships. A Graph Attention Network (GAT) is applied to automatically learn methylation patterns while simultaneously identifying informative subgraphs corresponding to DMRs. The proposed method is evaluated on 628 cell-free methylated DNA immunoprecipitation and high-throughput sequencing samples, achieving an AUC of 0.949 and outperforming previously reported methods. These results highlight the potential of graph neural networks for chromosome-level cfDNA analysis and cancer detection.

**Conditional Genetic Interaction Landscapes of Yeast WGD Paralogs Across Diverse Metabolic Environments**

**Authors List:**
**Mohammadreza Yasemi**, Concordia University, Canada
Brittany M Greco, Concordia University, Canada
Elena Kuzmin, Concordia University, McGill University, Canada

**Presentation Overview:**
Whole-genome duplication (WGD) in Saccharomyces cerevisiae (~100 million years ago) generated 551 duplicate gene pairs, and evidence suggests angiosperm-driven, sugar-rich fermentation niches contributed to their retention. Here, we profile 79 WGD paralog pairs (158 single and 79 double gene deletion mutants) characterized by sparse digenic and trigenic interactions in rich media and quantify fitness across 30 combinations of different carbon and nitrogen environments using automated time-lapse colony imaging. After correcting spatial, plate and batch effects, we compute area-under-the-curve growth metrics and use statistical models to identify gene–environment (G×E) and gene–gene–environment (G×G×E) effects that distinguish functional divergent and redundant paralogs. Preliminary analysis shows a substantial fraction of paralog mutants display condition-specific fitness defects. To score G×G×E, we adapted the ?-SGA trigenic interaction quantitative framework by treating the environment as the third axis. Conditional single- and double-mutant fitness defects are most consistently observed under maltose, galactose and raffinose-containing conditions, which are important for ripening and stress response. In contrast, glucose, fructose and sucrose conditions show fewer severe defects. Overall, the defect burden peaks in environments that require substantial metabolic rewiring for alternative carbon utilization, consistent with these niches being where paralog specialization and buffering are most strongly unmasked.

**A systematic assessment of the effect of data representation and machine learning methods on sample classification based on microbiota composition**

**Authors List:**
**Paniz Taghipour**, Memorial University of Newfoundland, Canada
Lourdes Pena-Castillo, Memorial University of Newfoundland, Canada

**Presentation Overview:**
Microbiota composition is associated with various phenotypes and diseases, making it an increasingly attractive target for diagnostic tools and therapeutic interventions. However, to achieve this, microbiome studies require reproducible, scalable workflows addressing compositional data constraints while balancing model interpretability with classification performance. We developed a Nextflow pipeline for 16S rRNA amplicon data analysis following best practices for microbiome analysis and evaluation of machine learning (ML) models. Our pipeline generates four commonly-used data transformations: genus-level relative abundances, presence–absence encodings, centered log-ratio transformations, and phylogenetic isometric log-ratio balances incorporating evolutionary relationships. Each of these data representations is used as input to a Random Forest classifier and a Neural Additive Model to generate models for sample classification. Each data transformation / ML method combination is evaluated using five independent datasets spanning a wide range of sample sizes ([50,1000]), hosts (human, dog, murre) and associated phenotype. We compare feature transformation methods and machine learning classifiers using the area under the precision–recall curve (AUPRC). Preliminary results indicate clear sample size dependencies and transformation-specific performance patterns. In particular, Neural Additive Models require substantially larger datasets than Random Forest to achieve equivalent predictive performance. Our results will provide evidence-based guidance for methods selection in microbiome studies.

**Dual curation of diploid genomes improves assembly quality**

**Authors List:**
**Erick Duarte**, The Rockefeller University, United States
Kirsty McCaffrey, The Rockefeller University, United States
Jack Medico, The Rockefeller University, United States
Giulio Formenti, The Rockefeller University, United States
Delphine Lariviere, The Rockefeller University, United States
Erich Jarvis, The Rockefeller University, United States

**Presentation Overview:**
Genome assembly algorithms face challenges such as high repetitiveness when piecing together sequence data from non-model genomes, often producing fragmented assemblies with errors. Correcting these misassemblies requires manual curation, including sequence breaks, joins, and reorientations at the contig and scaffold levels. which substantially improves assembly quality. Traditionally, curation has focused on a single haplotype or a pseudo-haploid representation of diploid genomes, potentially reducing haplotypic diversity and masking true variation. As fully phased assemblies become the standard, we present a “dual manual curation” methodology that involves simultaneous curation of both haplotypes of a diploid organism using a single Hi-C map. This approach minimizes structural variation errors, increases the accuracy of both haplotypes, improves the identification and curation of sex chromosomes, and yields two high-quality near-complete genomes. It also simplifies the resolution of misphasing events and streamlines the manual curation process. To support this approach, we have automated pre- and post-dual curation workflows in the Galaxy workspace. The automated workflows streamline the data processing, facilitate the data generation for curation decisions, and are less computationally demanding than the independent curation of both haplotypes. Dual curation has now become the main curation approach for the Vertebrate Genomes Project and Earth Biogenome Project.

**15:45-16:00**

**Afternoon Coffee Break**

**16:00-16:45**

**Fireside Chat - Do we really need another bioinformatics tool? To build or not to build.**

**16:45-17:00**

**Short Talk**

**Presentation Overview:**
Variant calling in real world somatic applications, is challenged by sample-specific artifacts, particularly in FFPE tissues where deamination, cross-linking, and coverage variability introduce systematic biases and drastically educe calling accuracy. This presentation describes the development of a unified secondary analysis framework within the DRAGEN™ suite that integrates multiple biology-aware variant detection tasks and incorporates machine learning–based error modelling to mitigate noise, demonstrating improved precision and sensitivity in both small variant and copy number detection across heterogeneous tumor samples.

**17:00-17:45**

**Success Circles**

## Friday, May 29th

**09:00-09:30**

**Plenary Talk - Quantum Computing for Bio-Discovery**

**Presentation Overview:**
I will discuss three problems in immunology, and bio-discovery more generally, that are being explored through quantum computing approaches on currently available noisy intermediate scale quantum (NISQ) hardware. Some of these explorations are preliminary while others are more advanced. In all of the these, the primary goal has been to extract some amount of reliable scientific meaning — what one might term

**09:30-09:45**

## TransXplorer: An Automated Discovery Engine for Identifying Understudied Biological Drivers Across Omics Modalities

**Authors List:**
**Varinder Verma**, University of Alberta, Canada
Eponine Oler, Department of Biological Sciences, University of Alberta, Canada
Hussain Syed, Department of Medicine, University of Alberta, Canada
Scott Han, Department of Biological Sciences, University of Alberta, Canada
Mark Berjanskii, Department of Biological Sciences, University of Alberta, Canada
Andrew L. Mason, Department of Medicine, University of Alberta, Canada
David S Wishart, Department of Biological Sciences, University of Alberta, Canada
Gane Ka-Shu Wong, Department of Biological Sciences, University of Alberta, Canada

**Presentation Overview:**
"Biologists often struggle to prioritize long lists of genes generated by high-throughput experiments, frequently focusing on well-characterized ""superstar"" genes while overlooking novel, understudied candidates. We present TransXplorer V2, an automated discovery engine that helps researchers move beyond the established literature to identify high-potential ""hidden drivers"" within any omics dataset, including transcriptomics, proteomics, and microbiome data. TransXplorer provides researchers with a prioritized candidate report that distinguishes between well-documented biological markers and ""Discovery"" candidates, genes that are mathematically central to the biological system but remain understudied in the literature. To achieve this, the software builds a comprehensive Neo4j-based Knowledge Graph (KG) digital map by integrating eight specialized databases. It evaluates each gene using a ""Five-Pillar"" scoring system that assesses existing evidence, network importance, regulation and research novelty. Finally, a team of specialized AI agents through LangGraph (Planner, Retriever, Synthesizer, and Critic nodes) synthesizes this data into a mechanistic report, using the “biological neighborhood” of a gene to propose its likely function even when direct literature is missing. This “Evidence-First” approach by combining mathematical network topology with adversarial hallucination checks by a Critic agent ensures that AI-generated interpretations are grounded in empirical data, providing a scalable, modality-agnostic tool for accelerating therapeutic discovery."

**09:45-10:00**

## Scalable Genomics Knowledge Discovery with Direct LLM–Database Integration

**Authors List:**
**David Yuan**, European Bioinformatics Institute, United Kingdom

**Presentation Overview:**
Large Language Models (LLMs) excel at interpreting unstructured text but remain limited in interrogating highly structured data natively. Existing attempts to bridge this gap—such as Model Context Protocols (MCP) and Knowledge Graphs (KG)—introduce substantial architectural complexity and scaling challenges, making them less suitable to the massive volumes of genomics data generated over the past four decades, including over 70 petabytes in the European Nucleotide Archive (ENA). We present a novel approach that integrates LLMs directly with relational databases, eliminating the need for external APIs or pre-materialized graph databases. In this framework, a knowledge graph is generated dynamically using a graph engine. A self-hosted LLM–backed AI agent then interrogates this graph using an AI-friendly Graph Query Language (GQL). As a proof of concept, we modeled antimicrobial resistance (AMR) phenotypes (over 1,500,000 entries) and genotypes (over 100,000 genomes and samples) and successfully queried them using free-text inputs via a web UI. This system enables Graph Retrieval-Augmented Generation (GraphRAG) directly over structured genomics data while remaining fully self-contained, reducing risks of model drift and data leakage. Importantly, it provides a highly intuitive interface for biologists, enabling exploratory queries and discovery beyond the constraints of pre-designed data portals.

**Evaluating the influence of annotation and statistical frameworks on pathway analysis results**

**Authors List:**
**Ritu Verma**, University of Guelph, Canada
Jeff Caswell, University of Guelph, Canada
Jeremy Simpson, University of Guelph, Canada
Anthony Mutsaers, University of Guelph, Canada
Sonja Fonfara, University of Guelph, Canada

**Presentation Overview:**
Hypertrophic cardiomyopathy (HCM) is a common acquired cardiac disease in cats and humans, yet its molecular mechanisms is not fully understood. RNA sequencing was performed on left ventricular and atrial tissue from healthy cats and cats with HCM, followed by pathway analysis using multiple tools, including DAVID, g:Profiler, Ingenuity Pathway Analysis (IPA), and Gene Set Enrichment Analysis (GSEA). Considerable variability was observed across pathway analysis approaches. Using default databases, DAVID and g:Profiler identified few mostly generalized biological processes, consistent with limited feline annotation. Using a custom feline background gene database for g:Profiler increased the number and diversity of detected pathways. IPA found numerous canonical pathways; filtering criteria influenced pathway rankings and statistical significance, and potential biologically relevant pathways were not always highly ranked. GSEA, using a custom feline pathway database, identified multiple biologically relevant Gene Ontology terms associated with cardiac structure, metabolism, and remodelling. Across methods, pathway detection and ranking were strongly influenced by database completeness, statistical framework, and filtering stringency, with increased thresholds relevantly reducing pathway detection. These results demonstrate the implication of database content and methodological choices when carrying out pathway analysis, which is of importance for transcriptomic studies in particular when analysing less-annotated genomes.

**NEXA: Agentic Network Exploration for Transparent Querying of Biomedical Knowledge Graphs**

**Authors List:**
**Mathieu Arru**, Département de médecine moléculaire, Université Laval & AxeEndo-Nephro, Centre de recherche du CHU de Québec, Canada
Mickaël Leclercq, Département de médecine moléculaire, Université Laval & AxeEndo-Nephro, Centre de recherche du CHU de Québec, Canada
Marie Pier Scott-Boyer, Département de médecine moléculaire, Université Laval & AxeEndo-Nephro, Centre de recherche du CHU de Québec, Canada
Arnaud Droit, Département de médecine moléculaire, Université Laval & AxeEndo-Nephro, Centre de recherche du CHU de Québec, Canada

**Presentation Overview:**
Introduction. Biomedical knowledge graphs (KGs) organize heterogeneous biological data into interconnected networks of genes, diseases, drugs, and pathways, supporting integrative biomedical analysis. However, use of KGs requires expertise in graph query languages, limiting accessibility. Large language models (LLMs) offer natural-language interfaces but often suffer from hallucinations and poor reproducibility. We introduce NEXA, a conversational agent designed for reliable and transparent natural-language querying of biomedical KGs. Methodology. The agent relies on a modular multi-step LLM pipeline that performs entity and relation detection, query generation, execution, and automatic correction. All answers are derived from executable KG queries, with intermediate steps exposed to the user for verification. A web interface enables interaction with multiple KGs, model selection, and exploration of retrieved subgraphs, supporting reproducibility and user oversight. Results. Benchmarking on an Alzheimer's disease knowledge graph shows up to 66% higher accuracy compared to baseline KGQA approaches. Complementary use-case studies illustrate effectiveness in answering complex biomedical queries, alongside gains in precision and interpretability. Conclusion. By tightly coupling LLM reasoning with verifiable KG queries, our approach enables trustworthy natural-language access to biomedical graphs. NEXA is model-agnostic, deployable locally, and suitable for privacy-sensitive biomedical environments, supporting transparent data exploration for research and clinical applications.

**Machine Learning Reveals an Extreme-Environment k-mer-based Watermark Imprinted Across Eukaryotic Genomes**

**Authors List:**
**Monireh Safari**, School of Computer Science, University of Waterloo, Canada
Joseph Butler, Department of Biology, University of Western Ontario, Canada
Gurjit S. Randhawa, School of Computer Science, University of Guelph, Canada
Victoria Kata, University of Western Ontario, Canada
Kathleen A. Hill, Department of Biology, University of Western Ontario, Canada
Lila Kari, School of Computer Science, University of Waterloo, Canada

**Presentation Overview:**
Extreme environments impose selective pressures that drive distinctive adaptations in extremophiles, organisms that thrive under conditions such as extreme temperature, salinity, or pH. Previous studies have shown detectable compositional signatures in prokaryotic genomes linked to extreme environments, enabling classification of bacteria and archaea by environmental niche using only DNA sequence. Whether comparable environmental signatures exist in eukaryotic extremophiles, however, remains unresolved, with important implications for evolutionary biology, biodiversity, and bioengineering. Here, we present a phylogeny-aware computational framework applied to 192 curated eukaryotic whole-genome assemblies spanning three kingdoms, Plantae, Fungi, and Animalia. Genomes are represented using k-mer frequency vectors (k = 3, 6, and 9) derived from randomly sampled genome fragments and analyzed using genus-level cross-validation. We identify a statistically significant, environment-associated component in eukaryotic nuclear DNA, achieving accuracies of 73.23% for extreme temperature, 80.21% for high salinity, and 62.67% for high pH environments (p 0.01). Using structural topic modeling (available as an open-source tool) with genus as a covariate, we identify co-occurring k-mer patterns whose environmental associations persist after controlling for phylogenetic structure. These results are the first evidence of k-mer–based environment-linked signatures in eukaryotic genomes and the first comprehensive, taxonomy-aware analysis distinguishing environment-linked sequence patterns from phylogenetic components.

**Multi-Omics Integration and Supervised Learning Identifies an Epithelial Signature for Radiotherapy Response in Colorectal Cancer**

**Authors List:**
Reuben Kumar, Department of Cancer and Genomic Sciences, College of Medicine and Health, University of Birmingham, UK, United Kingdom
Yujie He, Department of Cancer and Genomic Sciences, College of Medicine and Health, University of Birmingham, UK, United Kingdom
Jiarui Zhou, School of Biosciences and Centre for Environmental Research and Justice (CERJ), University of Birmingham, UK, United Kingdom
S-Cort Consortium, United Kingdom, multiple, United Kingdom
Andrew Beggs, Department of Cancer and Genomic Sciences, College of Medicine and Health, University of Birmingham, UK, United Kingdom
**Deena Gendoo**, Department of Cancer and Genomic Sciences, College of Medicine and Health, University of Birmingham, UK, United Kingdom

**Presentation Overview:**
Colorectal cancer (CRC) is the third most diagnosed cancer globally, and second leading cause of cancer deaths. Radiotherapy is a common treatment, but can demonstrate dangerous side effects and varying patient outcomes. Identifying biomarkers that can effectively predict patient responses to radiotherapy remains necessary for precise treatment of CRC patients. We channelled both unsupervised and supervised approaches to assess radiotherapy response for 233 patients of the S:CORT Consortium. Splitting the cohort, we first integrated matched RNA, CNA, mutation, and methylation profiles of 117 patients using multi-omics factor analysis (MOFA). We identified a radiotherapy signature of 101 biomarkers associated with patients who demonstrate a complete response to radiotherapy, and validated the signature using a random forest classifier on the internal validation dataset, and an independent testing cohort. Our signature effectively predicted treatment outcomes, achieving 89% accuracy with strong discriminatory performance (ROC_AUC = 0.85; PR_AUC = 0.71). Assessing human and murine scRNAseq datasets underscores that the signature is predominantly expressed in CRC epithelial cells, which underpin CRC heterogeneity and cellular diversity. Our identified signature enables pre-treatment identification of CRC patients that are unlikely to achieve a complete response to radiotherapy, thereby sparing these patients from unnecessary radiation exposure and off-target damage effects.

**10:00-10:45**

**Building Sustainable Funding for Bioinformatics**

**10:45-11:15**

**Morning Coffee Break**

**11:15-11:45**

## From Robust Cell Fates to Plastic States: Lineage Programs in Pediatric Brain Tumors

**Presentation Overview:**

Driver programs in pediatric brain tumors are tightly coupled to brain development. Our research leverages large, high resolution single cell references of the developing brain and in silico integration strategies to map patient tumors onto normal developmental trajectories and cellular states. This framework enables us to infer lineage of origin, differentiation state, and candidate vulnerabilities from multimodal tumor data. Here, I will focus on our recent work on PFA ependymoma, a lethal tumor of infants and toddlers with no major genetic drivers and few therapeutic options. There is a marked sex-difference in incidence and outcomes: boys are not only more likely to get the disease, but they also have a worse prognosis and shorter survival time. We find that tumor cells from male patients are less differentiated and more stem like than female tumor cells, and that the normal lineage from which these tumors arise mirrors this developmental delay. We also find that hormone signaling is responsible for this delay both in tumors and normal development, and that anti-androgen therapies represent a potential clinical avenue to target this deadly childhood cancer

**11:45-12:00**

## SIGMA : Targeted Cell Type Identification in scRNA-sequsing Gaussian Mixtures

**Authors List:**
Guillaume Lemaire, Centre de Recherche du Centre Hospitalier Universitaire de Québec-Université Laval, Québec, QC, Canada, Canada
Mickael Leclercq, Centre de Recherche du Centre Hospitalier Universitaire de Québec-Université Laval, Québec, QC, Canada, Canada
Éric Boilard, Centre de Recherche du Centre Hospitalier Universitaire de Québec-Université Laval, Québec, QC, Canada, Canada
Arnaud Droit, Centre de Recherche du Centre Hospitalier Universitaire de Québec-Université Laval, Québec, QC, Canada, Canada

**Presentation Overview:**
Single-cell RNA sequencing enables the characterization of cellular states. Extracting a specific cell type remains challenging. Clustering and manual annotation approaches depend on parameters, which complicates the detection of rare populations and increases result variability. We hypothesized that gene signatures would improve the identification of a target cell population. The goal was to develop a computational tool for precise extraction of a cell type of interest from raw data. We present SIGMA (Single-cell Identifier using Gaussian Mixtures Approaches), a bioinformatics pipeline that selects cells using Gaussian mixture models applied to target and exclusion gene signatures. The pipeline includes quality control and doublet removal, filters cells based on marker detection, assigns probabilities for each signature, and computes a final score. Across several public datasets, SIGMA consistently identified biologically coherent cell populations compared with existing approaches. It robustly recovered effector memory cytotoxic T lymphocytes across tissues and refined a COVID-19 megakaryocyte population by selecting cells with stronger and more consistent marker expression, while preserving known inflammatory subsets such as S100A8/A9-positive cells. SIGMA provides a reproducible framework for extracting high-confidence cellular populations from heterogeneous data, based on explicit and configurable rules, and outputs per-cell confidence scores.

**12:00-12:15**

## Epigenetic-State–Informed Polymer Modeling Reveals Tissue-Specific Principles of 3D Genome Folding

**Authors List:**
**Saman Bazmi**, University Health Network, Canada
Sushant Kumar, University Health Network, University of Toronto, Canada

**Presentation Overview:**

The spatial organization of the genome affects tissue-specific gene regulation by enabling long-range interactions between promoters and enhancers. Beyond static genomic architecture, chromosomal dynamics further influence gene expression through fluctuations in these long-range interactions. To explore the role of 3D genome folding in gene regulation, we integrated experimental Hi-C data, ChromHMM-derived epigenetic state annotations, and coarse-grained molecular dynamics simulations to analyze tissue-specific chromatin structures. We developed an optimized epigenetic-state-informed polymer modeling framework and used it to reconstruct tissue-specific chromatin folding landscapes, characterizing them with contact-probability scaling, compartment order parameters, radius of gyration, and other thermodynamic properties. Our analysis reveals significant tissue-dependent changes in compartment strength, spatial segregation, and local chromatin mixing that cannot be uncovered by ensemble-averaged Hi-C maps. We observed distinct patterns of chromatin compaction and radial redistribution of A/B compartments across genomic regions, reflecting functional specialization and context-dependent regulation. Collectively, these results establish a quantitative, scalable computational framework that links epigenetic state composition to emergent 3D genome structure across tissues, providing a basis for systematic investigation of genome reorganization in development and disease.

**Multiscale volume electron microscopy of the human liver maps vascular-cellular architecture, organelle dynamics and inter-organelle communication**

**Authors List:**
**Cheng Xing**, University of Toronto, Canada
Sonya MacParland, University of Toronto, Canada
Mei Zhen, University of Toronto, Canada
Gary Bader, University of Toronto, Canada

**Presentation Overview:**
The human liver depends on multiscale structural organization from vasculature to cells to organelles to perform its diverse metabolic functions. A unified three-dimensional view linking these hierarchical scales in intact human tissue would be useful for better understanding these levels and how they relate to each other. We present a high-resolution volume electron microscopy reconstruction of human periportal liver tissue acquired by serial block-face scanning electron microscopy. Using a multiscale deep learning approach, we performed automated segmentation across the entire volume, enabling comprehensive annotation of vasculature, cells, and organelles. Quantitative analysis of bile duct architecture revealed coordinated scaling between lumen geometry and cholangiocyte number and size. Sinusoidal capillary branches exhibited distinct structural profiles with differential endothelial coverage. Analysis of 35,790 complete mitochondria identified substantial morphological heterogeneity, with elongated mitochondria displaying preferential endoplasmic reticulum (ER) contacts at narrowing sites, a pattern consistent with ER-mediated fission or fusion. This multiscale reconstruction establishes an ultrastructural reference for the healthy human liver and provides a quantitative framework for investigating hepatic physiology and disease.

**Borderlands Science Automated : Large-scale reverberation through crowd augmentation**

**Authors List:**
**Parham Ghasemloo Gheidari**, McGill University, Canada
Kai-Hsiang Chang, McGill University, Canada
Bls Development Consortium, McGill University, Canada
Alexander Butyaev, McGill University, Canada
Roman Sarrazin-Gendron, Université du Québec à Montréal, Canada
Jérôme Waldispühl, McGill University, Canada

**Presentation Overview:**
Large-scale studies of the microbiome remain challenging due to several intertwined factors, such as reliance on host parameters and the dependence of methods on upstream analyses, such as multiple sequence alignment and phylogenetic tree construction. In 2020, we launched Borderlands Science, the first citizen science game embedded in an AAA video game, targeting multiple sequence alignment refinement. As a result, we gathered a dataset of more than 135 million human-curated annotations from 4 million players, which we showed improved microbial phylogeny. In parallel, we trained AI agents using behavioural cloning, successfully capturing and reproducing the best human solutions. Now, in our latest project, Borderlands Science Automated, we have automated the citizen science pipeline and integrated AI based on a new paradigm we call “crowd-augmentation,” empowering and upscaling the most important aspect of a citizen science project: the citizens. The final section of our pipeline encapsulates a recently developed beta-diversity method that can uncover biomarkers differentiating microbial communities. Overall, in our preliminary results, we have shown that our pipeline is capable of processing microbial data at a scale previously unattainable and with a quality that enables the discovery of potential biomarkers for several diseases and health statuses.

**Molecular Dynamics Simulations and Machine Learning For Therapeutic Strategies Discovery Against MSMP-Driven Cancers**

**Authors List:**
**Léopold Quitté**, Département de Médecine Moléculaire, Université Laval, Québec, QC, Canada., Canada
Arnaud Droit, Département de Médecine Moléculaire, Université Laval, Québec, QC, Canada., Canada
Mickael Leclercq, Centre de Recherche du CHU de Québec-Université Laval, Québec, QC, Canada., Canada
Gautier Moroy, Université Paris Cité, CNRS, INSERM, Unité de Biologie Fonctionnelle et Adaptative, F-75013 Paris, France., France

**Presentation Overview:**
The protein MicroSeMinoprotein Prostate associated (MSMP) is overexpressed in several types of disease, notably prostate, ovarian, and breast cancers. This overexpression is critical in tumors resistant to angiogenesis inhibitors therapies, where hypoxia-induced MSMP upregulation drives resistance. MSMP activates a cell receptor (CCR2) and its signaling G-protein, probably promoting tumor growth via a proliferation / differentiation / apoptosis pathway (MAP Kinase). To decipher this mechanism, we employed an advanced computational workflow. The MSMP-CCR2-G-protein complex was modeled using AlphaFold2 and analyzed through molecular dynamics simulations. To evaluate the specific impact of ligands, we compared this assembly to the well-known CCL2-CCR2-G-protein complex and the apo CCR2-G-protein system. These simulations were evaluated using gold-standard computational tools and in-house algorithms, specifically BioDiscML, an AI-based machine learning algorithm. Our results are consistent with experimental data, particularly regarding binding energies and receptor activation patterns. We demonstrated that while MSMP and CCL2 interact with CCR2 through distinct residues, they trigger similar activation mechanisms compared to the apo state. By pinpointing these specific molecular triggers, this study identifies key targets for drug discovery and provides a structural foundation for novel therapeutic strategies to overcome resistance in MSMP-related diseases.

**A Convolutional Deep Learning Approach to identify DNASequences for Gene Prediction**

**Authors List:**
**Jesus Antonio Motta**, Laval University, Canada

**Presentation Overview:**
We present a high?efficiency machine learning framework for identifying DNA sequences that encode human genes. Using GRCh38 as reference, we compiled training data from multiple curated genomic databases and translated nucleotide sequences into amino acid sequences to construct TF–IDF feature matrices. These representations were used to train a convolutional neural network (CNN) designed to learn discriminative patterns associated with coding regions. Training was performed across all 24 human chromosomes using approximately 36,000 genes and pseudogenes obtained from the HUGO Gene Nomenclature Committee (HGNC). Model performance was evaluated on 24 clinically relevant genes and their surrounding genomic regions. Metrics included precision, recall, F?score, accuracy, and ROC analysis. Our model demonstrates performance that exceeds expectations and achieves state?of?the?art accuracy for gene prediction, as confirmed through direct comparison with AUGUSTUS, our reference baseline. These results highlight the potential of deep learning–based sequence representations for robust gene identification across the human genome.

**12:15-13:00**

**Fireside Chat - Human VS Machine in Research Discoveries**

**14:30-14:35**

**CBH Award Introduction & Overview**

**14:35-15:00**

**CBH Award Distribution**

**15:00-16:00**

**Keynote - Decoding Early Life through Systems Biology – Distinct Immune Baseline Predicts Newborn Sepsis Risks**

**Presentation Overview:**
Each year, approximately three million newborns die from sepsis worldwide, with ~75% of all deaths under-five years occurring within the first week of life. Sepsis is a global concern, especially to young infants who are at highest risk of mortality. However, there lacks a quick and reliable way to identify those infants most at risk. To tackle this challenge, our group is applying systems biology approaches and supervised machine learning to identify newborn transcriptomic signatures that can predict sepsis at birth. Furthermore, using a segmented regression model coupled with elastic net, we established an early-life immune baseline that can predict newborn outcomes within the first month. Ultimately, our goal is to understand why newborns are highly susceptible to infections, and apply advanced omics strategies to develop real-world applications including vaccines, diagnostics and therapeutics.

**16:00-16:10**

**Let's Talk about IDEAS**

**16:10-16:30**

**Looking ahead at CBH**

**A-002: The complete telomere-to-telomere genome of the naked mole-rat provides insight into the genetics of healthy ageing and reproduction**

**Authors List:**
Zoe Clarke, University of Toronto, Canada
Jared Simpson, Ontario Institute for Cancer Research, Canada
Michael Wilson, The Hospital for Sick Children, Toronto, Canada
Gabriel Balmus, Cambridge University, United Kingdom
Remy Bruggman, University of Bern, Switzerland
Evan Eichler, Washington University, United States
Gary Lewin, Max Delbrück Center - Berlin, Germany
Daniel Mendez Aranda, Max Delbrück Center - Berlin, Germany
Daniel Hart, University of Pretoria, South Africa
Nigel Bennett, University of Pretoria, South Africa
Melissa Holmes, University of Toronto - Mississauga, Canada
Miguel Brieno-Enriquez, University of Pittsburgh, United States
**Dustin Sokolowski**, Ontario Institute for Cancer Research, Canada
Dongahn Yoo, Washington University, United States
Xinye Peng, University of Toronto - Mississauga, Canada
Mariela Faykoo-Martinez, University of Chicago, United States
Simon Monis, The Hospital for Sick Children, Canada
Sana Akhtar Alvi, The Hospital for Sick Children,

**Presentation Overview:**
The naked mole-rat (NMR; Heterocephalus glaber) is a eusocial rodent remarkable for a suite of traits including their exceptional longevity, cancer resistance, and unusual reproductive biology. Yet the genetic basis of these traits remains poorly understood. Here, we present a diploid, telomere-to-telomere genome assembly, annotated using long-read transcriptomic data from 16 adult tissues across both sexes. This assembly offers greatly improved resolution and annotation over previous versions, including a multiple multi-megabase long haplotype-specific segmental duplications that are found in wild populations. To investigate lineage-specific gene losses and duplications, we also generated a de novo long-read genome for the solitary, shorter-lived, and tumour-susceptible Cape dune mole-rat (Bathyergus suillus). Comparative analyses revealed NMR-specific gene losses linked to sperm dysfunction in the eusocial naked mole-rat and segmental duplications in pathways related to hypoxia tolerance, genome stability, and hormonal regulation. Together, these data establish a powerful genomic framework for understanding the molecular drivers of the naked mole-rat's extraordinary biology and provide a valuable resource for aging, cancer, reproductive, and evolutionary research.

## A-004: A Convolutional Deep Learning Approach to identify DNASequences for Gene Prediction

**Authors List:**
**Jesus Antonio Motta**, Laval University, Canada

**Presentation Overview:**
We present a high?efficiency machine learning framework for identifying DNA sequences that encode human genes. Using GRCh38 as reference, we compiled training data from multiple curated genomic databases and translated nucleotide sequences into amino acid sequences to construct TF–IDF feature matrices. These representations were used to train a convolutional neural network (CNN) designed to learn discriminative patterns associated with coding regions. Training was performed across all 24 human chromosomes using approximately 36,000 genes and pseudogenes obtained from the HUGO Gene Nomenclature Committee (HGNC). Model performance was evaluated on 24 clinically relevant genes and their surrounding genomic regions. Metrics included precision, recall, F?score, accuracy, and ROC analysis. Our model demonstrates performance that exceeds expectations and achieves state?of?the?art accuracy for gene prediction, as confirmed through direct comparison with AUGUSTUS, our reference baseline. These results highlight the potential of deep learning–based sequence representations for robust gene identification across the human genome.

## A-006: Integrating Single-Cell Coexpression Networks andMulti-omic Evidence for Drug Repurposing

**Authors List:**
**Rachel Edgar**, Ajmera Transplant Centre, Toronto General Research Institute, University Health Network, Canada
Sonya MacParland, University of Toronto (Immunology, Laboratory Medicine and Pathobiology); Ajmera Transplant Centre-UHN, Canada
Gary Bader, University of Toronto (Molecular Genetics, Computer Science, Donnelly Centre); PMCC-UHN; LTRI-Sinai Health; CIFAR, Canada

**Presentation Overview:**
Single-cell RNA-sequencing (scRNA-seq) has transformed our ability to resolve cellular heterogeneity in complex tissues and diseases. Greater understanding of tissue and cell type specific disease effects could lead to the development of targeted therapeutics. Drug repurposing, identifying new disease indications for existing drugs, is an efficient, cost-effective approach to therapeutic development. This has been successfully applied using bulk RNA-sequencing by linking disease transcriptional signatures to known drug targets.The generation of scRNA-seq maps for many diseases allows for the adaptation of these approaches to a cell type specific resolution. Here, we identify disease relevant gene modules by first constructing scRNA-seq coexpression networks from diseased tissue. Then within these cell type specific networks, gene modules are prioritized for disease relevance by integrating complementary data modalities, including bulk transcriptomics, genetic risk variants, and epigenetic associations. Finally disease-relevant modules are then tested for enrichment of known drug targets from DrugBank. We demonstrate the utility of this approach in the human liver and Primary Sclerosing Cholangitis (PSC), integrating cell type specific coexpression networks with PSC bulk transcriptomic and genetic variant data to uncover cell type specific therapeutic opportunities. In summary, this scalable framework enables scRNA-seq guided drug repurposing and more precise therapeutic prioritization.

## A-008: Understanding Treatment Response and Disease Outcomes in Pedaitric Lupus Using Whole Blood RNA-sequencing

**Authors List:**
**Isha Datar**, Molecular Genetics, University of Toronto, Canada
Huayun Hou, The Hospital for Sick Children, Canada
Reagan Reid, Institute of Medical Sciences, University of Toronto, Canada
Daniela Dominguez, Pediatric Rheumatology, University of Toronto, Canada
Lauren Erdman, School of Medicine, University of Cincinnati, United States
Eleanor Pullenayegum, Dalla Lana School of Public Health, University of Toronto, Canada
Linda Hiraki, The Hospital for Sick Children, Canada
Michael Wilson, The Hospital for Sick Children, Canada

**Presentation Overview:**
Systemic lupus erythematosus (SLE) is a chronic autoimmune disease affecting multiple organs. Pediatric lupus makes up 20% of all lupus cases but has higher morbidity and mortality than adult-onset lupus. Front-line treatment relies on corticosteroids which act systemically, modulating gene expression to reduce inflammation. However, disease management remains a challenge due to diverse clinical manifestations and variable treatment response. Existing gene expression profiling studies have overlooked patient treatment history, recruiting patients several years after disease onset when their disease state has been modified by disease associated complications and medication. To understand heterogeneity in disease manifestation and treatment response there is a need to study treatment naïve patients and systemically follow these patients as their immune cell activity changes with treatment. Along with our collaborators, we leveraged the SickKids Lupus Clinic that enrols lupus patients at diagnosis and follows them across the treatment course. We generated whole blood bulk RNA sequencing data from 74 pediatric lupus patients before treatment, 6 months and 15 months post-treatment. I aim to use this longitudinal gene expression data to understand variability in treatment response. To understand cell type specific treatment effects, I applied BayesPrism cellular deconvolution to generate immune cell specific gene expression profiles. I identified pan-cell type and unique cell type genes and pathway level changes after treatment. I aim to correlate these changes to clinical disease manifestations and identify biomarker-like gene profiles in a cell type specific manner that predict favorable response to corticosteroid treatment.

## A-010: A Unified Heterogeneous Graph Transformer for TCR-Epitope Binding and Tumor-Reactive T Cell Identification

**Authors List:**
Yunsheng Chen, Algoma University, Canada
**Vanessa Giuliano**, Algoma University, Canada
Yan Yan, University of Guelph, Canada
Wenjun Lin, Algoma University, Canada
Ping Luo, Algoma University, Canada

**Presentation Overview:**

T cell receptor (TCR) recognition is central to adaptive immunity and cancer immunotherapy. Although computational approaches have been developed for predicting TCR-epitope binding, existing methods typically focus on isolated tasks and lack a unified framework capable of addressing both traditional binding prediction and emerging single-cell tumor reactivity classification. Bridging these tasks is essential for translating sequence-level insights into functional immune responses. We present a heterogeneous graph transformer framework that jointly addresses TCR-epitope binding prediction and tumor-reactive T cell identification. The model integrates sequence embeddings with a multi-relational graph structure to capture biological relationships among TCRs, epitopes and T cells. Advanced attention mechanisms are employed to model both local and global dependencies, including degree centrality encoding to represent hub nodes and Laplacian positional encoding to preserve global graph context. Cross-attention mechanisms further facilitate information exchange, enabling unified learning across molecular and cellular levels. Our model achieves AUCs of 0.927 and 0.955 for TCR-epitope binding prediction on the IEDB and VDJdb datasets, respectively, and an AUC of 1.00 for tumor-reactive T cell classification using human gastrointestinal cancer single-cell data. Collectively, these results demonstrate the performance and generalizability of our unified graph-based framework, highlighting its potential to advance TCR modeling and immunotherapy research.

## A-012: Isoform-Switch Biomarkers Predict Drug Response andPrognosis in Acute Myeloid Leukemia

**Authors List:**
**Stefan Wallin**, University of Ottawa, Canada
Arvind Mer, University of Ottawa, Canada

**Presentation Overview:**
Acute myeloid leukemia (AML) remains a highly lethal malignancy, in part due to substantial molecular heterogeneity and the development of therapy resistance. While alternative splicing is increasingly recognized as a driver of cancer progression, most AML studies rely on gene-level analyses that fail to capture functionally distinct isoforms. Here, we systematically investigate isoform-level expression and isoform-switching events to identify biomarkers of prognosis and drug response in AML. Transcript-level RNA-seq data from four independent AML cohorts (>650 patients) were integrated with pharmacogenomic datasets spanning 26 compounds and 14 drug classes. Isoform-switch ratios were tested for associations with overall survival using multivariate Cox regression and with drug response using Spearman correlations, with results combined by random-effects meta-analysis. We identified 468 robust isoform-switches associated with survival and drug resistance, particularly in MAPK- and apoptosis-related pathways, and prioritized 15 high-confidence biomarkers from known cancer genes. An isoform-level protein-protein interaction network revealed how splicing alters molecular connectivity, including an IKBKG isoform switch that disrupted the ATM-NEMO interaction and was linked to MEK inhibitor resistance and poor survival. These findings highlight isoform-level analysis as a powerful framework for identifying clinically relevant biomarkers and therapeutic vulnerabilities in AML.

## A-014: Significance of Receptor Membrane Embedding for AccurateMolecular Dynamics Simulations

**Authors List:**
**Jesus Castano**, Département de Biomédecine Vétérinaire, Faculté de Médecine Vétérinaire, Université de Montréal, Canada
Francis Beaudry, Département de Biomédecine Vétérinaire, Faculté de Médecine Vétérinaire, Université de Montréal, Canada

**Presentation Overview:**
Computational simulations offer a unique platform capable of identifying molecular interactions at the atomic level, uncovering new mechanistic principles that could help advance drug design strategies. However, since simulations rely on approximations of complex biological systems, simulation artifacts can be generated when using incorrect approaches. In this work, we compared the effect of membrane embedding for transmembrane receptor-ligand interactions in TRPV-like proteins – a common target for analgesic drugs. Molecular docking was carried out using AutoDockFR, while membrane embedding was performed in the CHARMM-GUI platform, using POPC as a model lipid bilayer. For comparative analysis, we ran two Molecular Dynamics (MD) simulations using Gromacs 2024.4, with and without a membrane. The systems were energy-minimized, equilibrated, and subjected to 100-ns simulations, while the corresponding MD trajectories were analyzed using the MDAnalysis package. The results revealed that without a membrane, critical ligand-receptor interactions such as hydrogen bonds and polar interactions were severely reduced (~100% in some cases) or even disappeared, affecting conformational sampling, and free energy landscapes. This could result in biological implications for drug design and hypothesis generation, which could mislead research efforts for novel analgesics. Therefore, our study highlights the importance of membrane embedding during ligand-receptor calculations for transmembrane systems.

## A-016: Systematic Transcriptomic Biases Distinguish Single-Celland Single-Nucleus RNA Sequencing

**Authors List:**
**Camdyn Holmes**, Department of Biochemistry, Schulich School of Medicine & Dentistry, Western University, Canada
Tallulah Andrews, Department of Biochemistry, Schulich School of Medicine & Dentistry; Department of Computer Science, Western University, Canada

**Presentation Overview:**
Single-cell RNA sequencing (scRNA-seq) has revolutionized transcriptomics but exhibits systematic capture biases, enriching easily dissociated immune cells while under-representing adhesive populations such as epithelial cells and hepatocytes. Single-nucleus RNA sequencing (snRNA-seq) improves recovery of these cell types but is biased against immune cells. Integration of scRNA-seq and snRNA-seq is complicated by many genes exhibiting technical and biological differences in their expression between these protocols. We characterized potential biological features that drive these differences between 9 matched scRNA-seq and snRNA-seq datasets. The largest effect was due to preferential enrichment of long genes in snRNA-seq, consistent with prolonged transcriptional time and nuclear residency for longer genes. Mitochondrial and ribosomal genes were enriched in scRNA-seq, reflecting cytoplasmic RNA capture. Most scRNA-seq datasets showed upregulation of stress response genes, indicating dissociation-associated expression alteration. In contrast, technical factors such as internal priming had little impact. Membrane and cell-surface marker genes were underrepresented in snRNA-seq, likely reflecting reduced capture of cytoplasmic mRNA. Many of these genes are key markers of immune populations, thus further preventing characterization of immune cell types in snRNA-seq. Future work should focus on correcting these biases to improve integration, and/or novel methods to collect scRNA-seq and snRNA-seq more cost-effectively.

## A-018: It is High Time for Rapid Generative Artificial Intelligence Learning in Oncology

**Authors List:**
**Youcef Derbal**, Toronto Metropolitan University, Canada

**Presentation Overview:**
Therapeutic resistance remains one of the most critical barriers to achieving durable cancer control or cure. Although the range of available treatments continues to expand, with advances in targeted therapies, immunotherapy, radiotherapy, chemotherapy, and their combinations, resistance almost inevitably emerges due to the complex evolutionary dynamics of cancer and its inter-patient heterogeneity. Overcoming or delaying resistance requires treatment strategies that can learn from and adapt to the tumor’s eco-evolutionary dynamics, informed by the patient’s clinical, molecular, and radiomic characteristics. Generative artificial intelligence (GenAI) provides a powerful framework for rapid learning in oncology by integrating continuously acquired multimodal patient data to guide personalized, adaptive cancer therapies. Realizing this potential requires broader access to cancer data, particularly data generated through publicly funded healthcare systems. Such data should be leveraged for public benefit, enabling accelerated discovery and more effective treatments across the population. To support this vision, public health institutions must be legally and financially empowered to reduce barriers to cancer data access and to accelerate clinical trials of GenAI-driven treatment modalities, ensuring timely validation and responsible integration of these technologies into cancer care.

## A-020: A Deep Learning Model for Predicting Disease ProgressionLevels Using Single-Cell RNA Sequencing Data

**Authors List:**
**Jordan Babadoudou**, Montreal University, Canada
Yoshiaki Tanaka, Maisonneuve-Rosemont Hospital Research Center, Canada

**Presentation Overview:**
The brain's complexity means that cellular dysfunctions often manifest as shared molecular mechanisms across diverse neurological disorders. For instance, aging is a common risk factor for both Alzheimer's and Parkinson's diseases. However, identifying these overlapping pathological signatures remains a major challenge. Here, we leverage weakly-supervised deep learning model that is designed to classify individual cells based on their risk trajectories and segregate cells into diseased and healthy states. This model is trained by a comprehensive collection of single-cell RNA sequencing dataset comprising over 10 million cells from 750 patients across glioma, Alzheimer's, autism, epilepsy, and multiple sclerosis, and identifies key shared pathological mechanisms across different diseases. Overall, this research provides a framework for understanding how different brain diseases converge at a cellular level. Ultimately, this work aims to facilitate the development of multi-disease therapeutic strategies, potentially improving treatment outcomes for multiple brain pathologies simultaneously.

## A-022: HiMaLAYAS: enrichment-based annotation of hierarchically clustered matrices

**Authors List:**
**Ira Horecka**, University of Toronto: Department of Molecular Genetics, Canada
Hannes Röst, University of Toronto: Department of Molecular Genetics, Department of Computer Science, Canada

**Presentation Overview:**
Hierarchical clustering organizes high-dimensional biological matrices and is commonly used for visualization rather than statistical inference. Most enrichment-based analyses of hierarchically clustered matrices are confined to gene expression data and fixed workflows. We introduce Hierarchical Matrix Layout and Annotation Software (HiMaLAYAS), a framework for post hoc enrichment-based annotation of such matrices. HiMaLAYAS treats clusters as statistical units, tests annotation enrichment, and renders significant annotations alongside the matrix. We apply HiMaLAYAS to

## A-024: Docking and Molecular Dynamics Insights into Dopamine andL-DOPA Recognition by Human and C. elegans DopamineReceptors

**Authors List:**
**Lusine Mkrtchyan**, Département de Biomédecine Vétérinaire, Faculté de Médecine Vétérinaire, Université de Montréal, Québec, Canada, Canada
Francis Beaudry, Département de Biomédecine Vétérinaire, Faculté de Médecine Vétérinaire, Université de Montréal, Québec, Canada, Canada
Jesus D. Castaño, Département de Biomédecine Vétérinaire, Faculté de Médecine Vétérinaire, Université de Montréal, Québec, Canada, Canada

**Presentation Overview:**
Parkinson's disease (PD) is characterized by dopaminergic neuron degeneration and remains primarily managed with L-DOPA therapies that alleviate symptoms without halting progression. Understanding dopamine receptor recognition across model organisms is essential for translational research. We performed comparative computational analysis of dopamine and L-DOPA interactions with human and Caenorhabditis elegans D1- and D2-like dopamine receptors to evaluate evolutionary conservation of dopaminergic signaling mechanisms. AlphaFold-derived models were validated before simulation studies. Using flexible molecular docking and 100-ns molecular dynamics simulations, we demonstrate that both ligands engage conserved orthosteric pockets across species and receptor subtypes despite ~600 million years of divergence. Dopamine and L-DOPA exhibited comparable binding affinities, particularly for D2-type receptors, with stable positioning and preserved GPCR architecture. Principal component and free energy landscape analyses revealed ligand-dependent dynamics: dopamine favored focused low-energy states, while L-DOPA induced broader conformational ensembles, especially in human receptors. Interaction profiling identified a conserved five-residue recognition core (Val/Ile3.33, Ser5.43, Phe5.47, Phe6.51, His6.55/Asn6.55). Receptors achieved high-affinity binding through modular strategies including canonical Asp3.32 ionic anchoring, alternative Lys-mediated stabilization in human D1 receptors, and species-specific aromatic interactions in C. elegans. These results demonstrate deep structural conservation and support C. elegans as a mechanistically relevant PD model.

## A-028: A Machine Learning Approach to Identify Determinants ofsnoRNA–Protein Interactions

**Authors List:**
**Zakaria Meliane**, Université de Sherbrooke, Canada
Michelle Scott, Université de Sherbrooke, Canada

**Presentation Overview:**
Small nucleolar RNAs (snoRNAs) are highly conserved non-coding RNAs crucial for ribosome biogenesis through interactions with core binding proteins. C/D box snoRNAs, interact with NOP58, NOP56, SNU13, and the methyltransferase FBL. Multiple studies have revealed a much broader functional landscape for snoRNAs, including roles in which they interact with other proteins. Additionally, large-scale RNA-protein binding studies indicate that not all snoRNAs interact with the snoRNA core binding proteins. Furthermore, the molecular features underlying these interactions remain poorly understood. To elucidate these questions and gain insight into the snoRNA-core binding protein interaction determinants and improve snoRNA functional classification, we use a machine learning approach to identify features contributing to snoRNA-protein complex formation. The features considered in this study include the sequence of the C, C', D and D' box motifs, their absolute and relative positioning , the presence of k-turn structural element, and snoRNA length. We aim to classify whether a given C/D box snoRNA interacts with a protein of interest and to identify which of these features most strongly contribute to complex formation. Collectively, this work shows that our understanding of snoRNAs remains limited, and that a deeper characterization of their fundamental properties will improve insights into their biological functions.

**A-030: Multi-Omics Reveals Distinct Molecular Signatures of Cardiac Hypertrophy in Hypertensive versus Genetic Heart Disease.**

**Authors List:**
**Mobin Khoramjoo**, University of Alberta, Faculty of Medicine and Dentistry, Department of Physiology, Canada
Huachen Chen, University of Alberta, Faculty of Medicine and Dentistry, Department of Medicine, Canada
Anran Zhang, University of Alberta, Faculty of Medicine and Dentistry, Department of Physiology, Canada
Zamaneh Kassiri, University of Alberta, Faculty of Medicine and Dentistry, Department of Physiology, Canada
Gavin Oudit, University of Alberta, Faculty of Medicine and Dentistry, Department of Medicine, Canada

**Presentation Overview:**
Left ventricular hypertrophy (LVH) represents a convergent phenotype driven by distinct etiologies: pressure overload of hypertension (HTN) or the genetic mutations in hypertrophic cardiomyopathy (HCM). To characterize myocardial hypertrophy in these etiologies, we performed histological evaluations and multi-omics profiling, including WGS, bulk RNA-seq, and metabolomics on left ventricle of explanted human hearts from 13 non-failing controls (NFC), 19 HTN-LVH, and 19 HCM age and sex matched subjects. We leveraged bioinformatic and ML approaches to detect etiology-specific biomarkers and pathogenic mechanisms. WGS confirmed non-genetic hypertrophy in HTN-LVH. Transcriptomics revealed more DEGs in HCM vs NFC (4640) than in HTN-LVH vs NFC (389); however, GSEA identified more dysregulated GO pathways in HTN-LVH, indicating earlier-stage remodelling with broader, while less pronounced transcriptional changes. Many pathways were etiology-specific, suggesting distinct molecular routes to hypertrophy. Metabolomics revealed disruptions in glucose and fatty acid oxidation exclusively in HCM, reflecting advanced metabolic remodelling and energy exhaustion of severe disease. Multi-omics integration (MOFA) identified latent factors associated with hypertrophy severity across molecular layers, while elastic net regression revealed ECM-related genes, including ADAMTS6, strongly correlated with myocardial fibrosis. Histological analysis revealed greater cardiomyocyte hypertrophy in HTN-LVH than in HCM, while fibrosis was more prominent in HCM, indicating distinct pathological mechanisms underlying these etiologies. Multi-omics profiling distinguished HTN-LVH from HCM, enabling precision medicine by identifying differential dysregulated pathways and novel therapeutic targets for heart disease.

## A-032: Fostering Global Collaboration in Bioinformatics with theSociété Française de Bioinformatique

**Authors List:**
**Charles Lecellier**, CNRS, France

**Presentation Overview:**
The Société Française de Bioinformatique (SFBI) is France's leading society, uniting biology, computer science, mathematics, and physics to advance bioinformatics. As a volunteer-driven organization, SFBI promotes research, education, and innovation through initiatives such as the organisation of the annual JOBIM conference, professional training, and networking opportunities. It also supports career development by disseminating job openings and fostering discussions on bioinformatics education. SFBI is committed to global collaboration, as exemplified by the co-founding of the West African Network of Bioinformaticians and System Administrators (RABIAS) in December 2025, which strengthens regional research capacity in health and agriculture. At JOBIM 2026, SFBI will host a symposium titled 'International Interactions in Bioinformatics: Sequencing Technologies for Health and Agriculture', providing a platform to explore the challenges and opportunities of cross-border partnerships with researchers from West Africa and South America. Given their shared missions and goals, SFBI would like to explore new partnerships with the Canadian Bioinformatics Hub (CBH). By combining resources and expertise, SFBI and CBH could amplify their impact through joint workshops, conferences, and mentorship programs, and create unique opportunities for researchers from Canada and France. These new collaborations have the potential to catalyze innovation and strengthen connections within the global bioinformatics community.

## A-034: A Bioinformatics Workflow and A Comparative Profile of TwoExtracellular RNAs Across Different Blood Collection Tubes

**Authors List:**
**Enaam Merchant**, Norgen Biotek Corp., Canada
Hinal Zala, Norgen Biotek Corp, Canada
Alex Chauhan, Norgen Biotek Corp, Canada
Simone Yamasaki, Norgen Biotek Corp, Canada
Songsong Geng, Norgen Biotek Corp, Canada
Mohamed El-Mogy, Norgen Biotek Corp, Canada
Taha Haj-Ahmad, Norgen Biotek Corp, Canada
Yousef Haj-Ahmad, Norgen Biotek Corp, Canada

**Presentation Overview:**
The widespread use of BCTs (blood collection tubes) across clinical and research settings raises questions regarding the biological components captured, sources of variability, and analytical outcomes. An area that remains understudied is the comparative characteristics of extracellular RNA (exRNA). To explore these concepts, human exosomal RNA and cell-free RNA (cfRNA) were isolated from three chemically distinct BCTs and sequenced with identical protocols to generate Illumina paired-end reads. Primary data processing assessed sequencing quality and initial genome & transcriptome mapping performances. Additional analyses studied genomic classification, gene normalization & distribution, RNA splicing quantification, and RNA quality metrics. Statistical significance and effect size were measured using the Wilcoxon Rank-Sum Test. The exRNA results indicate that cfRNA exhibited higher mapping rates (17.92% higher for the genome) and greater transcriptomic diversity (approximately 12,487 more genes) than exosomal RNA. However, comparative feature analysis also revealed biological overlap, and differences characterized by small statistical effect sizes. This pattern of low-effect-size differences between the two exRNAs was also closely observed across the individual BCT type. This study establishes a foundational bioinformatics workflow to measure the subtle variations and expression profiles for quantifying exRNA differences arising from biological compartments and BCT choices.

## A-036: Predicting Cytogenetic Abnormalities in Multiple Myelomafrom Gene Expression Data using Machine Learning

**Authors List:**
**Sarah Ryan**, University of Ottawa, Canada
Ben Patrick, University of Ottawa, Canada
Braeden Walker, University of Ottawa, Canada
Arvind Mer, University of Ottawa, Canada

**Presentation Overview:**
Multiple myeloma is an incurable blood cancer, defined by genetic abnormalities which convey important prognostic and treatment implications. Standard-of-care testing uses fluorescence in-situ hybridization (FISH), which only captures limited information on several cytogenetic abnormalities. Conversely, RNA sequencing (RNA-seq) provides a wider, more nuanced breadth of tumour information. In this study, we investigate whether FISH information can be extracted from RNA-seq data. We developed machine learning models to predict clinically relevant cytogenetic abnormalities using the MMRF-CoMMpass study data (NCT01454297). Using 661 samples, we trained Random Forest (RF) and XGBoost (XBG) classifiers to predict five lesions: del(17p13), gain(1q21), t(11;14), t(14;16), t(4;14). Features were biologically informed, derived from genes on involved chromosomes and relevant hallmark pathways. Models underwent 4-fold cross-validation with hyperparameter tuning, custom weighting, and threshold optimization. Models achieved comparable performance (RF: average F1 score 0.91±0.06, subset accuracy 86.5%; XGB: F1 0.91±0.08, subset accuracy 87.6%). Lesion-specific F1 scores ranged from 0.79 (del(17p13)) to 0.98 (t(11;14)). Future work involves external validation in independent cohorts and across different transcriptomic profiling platforms to assess generalizability and potential for clinical implementation.

## A-038: On-Premise Compute Setup for Air-Gapped Inference:Bootstraps and Shoestrings.

**Authors List:**
**Christopher W. V. Hogue**, unaffiliated, Canada

**Presentation Overview:**
A small, self-hosted, cloud-native, GPU based compute cluster is described. It provides physical isolation and operational autonomy while ensuring localized data sovereignty. The intended system workload is mixed, multi-agent, image and language model inference, using quantized open source models. It is a 6-node bare metal Kubernetes cluster that uses small, server grade, power efficient motherboards and workstation GPUs. Rapid initial configuration iterations are achieved by splitting it into two 3 node clusters - one operates while the other is being rebuilt. The system provides general and specialized inference services with open access LLMs without ever sending data outside the facility. In addition to data privacy, it eliminates token costs associated with cloud based inference. I cover the system setup dependency matrix, networking, multiuser authentication, certificate management, top-of rack components, bare metal node OS provisioning, GPU operators, IP address handouts, cluster ingress, container and model caching, cluster management and systemwide monitoring. Multiuser authentication, authorization, ssh-certificates, and dedicated service accounts allow internal AI agents to access the system with well-defined constraints. A product of nearly 30 years of experience and iterations, it is a mashup of a nimble academic compute lab and a secure, air-gapped, telco cloud edge installation.

## A-040: Optimizing scRNAseq analysis for atlas-level projects.

**Authors List:**
**Tallulah Andrews**, Department of Biochemistry, University of Western Ontario, Canada

**Presentation Overview:**
Single-cell RNAseq is now a standard approach in biomedical research, with thousands of datasets and billions of cells sequenced. Increasingly, researchers are combining these datasets to generate atlases of both healthy and diseased tissues. However, technical biases, and poor scalability limit our ability to combine datasets and perform cross-condition comparisons. Here, we address these challenges by characterizing the unique technical biases observed between 3' and 5' data; and the biological biases between single-cell and single-nucleus RNAseq. In addition, we introduce M3Drop2, and optimized negative binomial model for feature selection, data normalization, gene-gene correlations, and batch effect correction on atlas scale datasets. We highlight computational strategies to improve algorithm scalability, including data-chunking and GPU integration. We show that M3Drop2 outperforms sctransform for integrating 3' and 5' data. In addition, we demonstrate its scalability by integrating two existing mouse brain atlases to generate the largest map of mouse neuronal diversity available, and identifying co-expression patterns of neurotransmitters across >100 neuron cell-types. Together, these results demonstrate the importance and practicality of scaling data analysis tools to millions of cells to empower novel discoveries through combining existing single-cell datasets.

## A-042: SIGMA : Targeted Cell Type Identification in scRNA-sequsing Gaussian Mixtures

**Authors List:**
**Guillaume Lemaire**, Centre de Recherche du Centre Hospitalier Universitaire de Québec-Université Laval, Québec, QC, Canada, Canada
Mickael Leclercq, Centre de Recherche du Centre Hospitalier Universitaire de Québec-Université Laval, Québec, QC, Canada, Canada
Éric Boilard, Centre de Recherche du Centre Hospitalier Universitaire de Québec-Université Laval, Québec, QC, Canada, Canada
Arnaud Droit, Centre de Recherche du Centre Hospitalier Universitaire de Québec-Université Laval, Québec, QC, Canada, Canada

**Presentation Overview:**
Single-cell RNA sequencing enables the characterization of cellular states. Extracting a specific cell type remains challenging. Clustering and manual annotation approaches depend on parameters, which complicates the detection of rare populations and increases result variability. We hypothesized that gene signatures would improve the identification of a target cell population. The goal was to develop a computational tool for precise extraction of a cell type of interest from raw data. We present SIGMA (Single-cell Identifier using Gaussian Mixtures Approaches), a bioinformatics pipeline that selects cells using Gaussian mixture models applied to target and exclusion gene signatures. The pipeline includes quality control and doublet removal, filters cells based on marker detection, assigns probabilities for each signature, and computes a final score. Across several public datasets, SIGMA consistently identified biologically coherent cell populations compared with existing approaches. It robustly recovered effector memory cytotoxic T lymphocytes across tissues and refined a COVID-19 megakaryocyte population by selecting cells with stronger and more consistent marker expression, while preserving known inflammatory subsets such as S100A8/A9-positive cells. SIGMA provides a reproducible framework for extracting high-confidence cellular populations from heterogeneous data, based on explicit and configurable rules, and outputs per-cell confidence scores.

## A-044: Integrated Transcriptomic and Epigenomic Profiling of RYR1-Related Disorders

**Authors List:**
**Marjorie Labrecque**, CHUM Research Center, Montreal, Quebec, Canada, Canada
Jean-Denis Brisson, Neuromuscular Diseases Clinic, Saguenay-Lac-Saint-Jean,Quebec, Canada, Canada
Clément Morin, Neuromuscular Diseases Clinic, Saguenay-Lac-Saint-Jean,Quebec, Canada, Canada
Jean Mathieu, Neuromuscular Diseases Clinic, Saguenay-Lac-Saint-Jean,Quebec, Canada, Canada
Alex Parker, CHUM Research Center, Montreal, Quebec, Canada, Canada
Cam-Tu Emilie Nguyen, Sainte-Justine University Hospital Center (CHSJ), Montreal,Quebec, Canada, Canada
Martine Tetreault, CHUM Research Center, Montreal, Quebec, Canada, Canada
Jennifer Hautecloque, CHUM Research Center, Montreal, Quebec, Canada, Canada
Adrien Rihoux, CHUM Research Center, Montreal, Quebec, Canada, Canada

**Presentation Overview:**
RYR1-related disorders (RYR1-RD) are a subtype of myopathies marked by clinically diverse phenotypes. The RYR1 gene plays a pivotal role in skeletal muscle contraction. Our team identified a novel, previously uncharacterized pathogenic missense variant in RYR1, p.E176K, in 28 individuals from six unrelated families exhibiting inter- and intra-familial phenotypic variability. We hypothesize that the signatures obtained through multi-omics will contribute to a better understanding of the observed heterogeneity. Muscle biopsies from four RYR1-RD patients carrying the heterozygous p.E176K variant and one control were extracted for sequencing. Differential gene expression, alternative splicing, and alternative polyadenylation were obtained from RNA-seq data, while differentially methylated regions were identified from whole genome bisulfite sequencing. Gene sets enrichment analysis was done in ShinyGO. Our analysis revealed 283 differentially expressed genes, 196 alternatively spliced genes, 1140 alternatively polyadenylated genes and 1515 differentially methylated regions. Gene set enrichment analysis highlighted six biological processes related to muscle contraction that were consistently dysregulated across all omics layers. These results demonstrate extensive dysregulation of transcriptomic and epigenomic profiles in RYR1-RD and provide insight into the molecular basis of phenotypic variability. As a next step, we will sequence additional samples and correlate these signatures with disease severity across affected individuals.

## A-046: D2R-HeDS: A Health Data Science Platform to Support RNAResearch

**Authors List:**
**Senthilkumar Kailasam**, D2R-HEDS McGill University, Canada
Eric Audemard, McGill University, Canada
Sebastian Ballesteros, McGill University, Canada
Guillaume Bourque, McGill University, Canada
Mathieu Bourgey, McGill University, Canada
Maryam Youssef, McGill University, Canada

**Presentation Overview:**
As part of a university-based DNA to RNA (D2R) initiative, the Health Data Science (HeDS) platform provides end-to-end infrastructure and expertise that accelerate data-driven discovery across genomics, transcriptomics, and RNA therapeutics. HeDS operates at the intersection of research computing, data engineering, and applied bioinformatics/AI, enabling projects to move from raw data to reproducible results while maintaining robust governance and data sovereignty. HeDS delivers four integrated capability areas. (1) Data management, curation, and interoperability: research data management guidelines and best practices help researchers harmonize metadata capture, use controlled vocabularies, and onboard datasets into a shared metadata catalogue to improve discoverability, collaboration, and reuse. (2) Analytical and methodological support: hands-on bioinformatics, biostatistics, and ML/AI support enhances RNA-focused research. (3) Secure data environments and compute: compliant workspaces and scalable pipelines on institutional and national infrastructure enable reproducible execution using containerized workflows and versioned analysis environments. (4) Translation and enablement: training workshops, reusable templates, consultation, and student mentoring increase data science capacity in the RNA research community. Together, HeDS provides a scalable foundation for a rigorous, interoperable, and collaborative data science environment, strengthening research quality, accelerating discovery, and increasing the impact of the initiative and partner programs.

## A-048: Synthetic Generation of Breast Cancer Transcriptomic Profiles Using a Denoising Diffusion Probabilistic Model for Precision Oncology

**Authors List:**
**Alfredo Varela-Vega**, Department of Biochemistry, Schulich School of Medicine & Dentistry, Western University, London, Ontario, Canada., Canada
Pingzhao Hu, Departments of Biochemistry and Computer Science, Western University, London, Ontario, Canada., Canada

**Presentation Overview:**
Breast cancer remains the most prevalent cancer among women worldwide, even in the era of precision oncology and artificial intelligence. Recent advances have enabled the development of algorithms capable of learning general patterns from large genomic datasets which can be adapted to personalized diagnosis, biomarker discovery, and survival analysis. However, these models are often limited by the availability, size, and diversity of training datasets. This can lead to overrepresentation of genomes and expression profiles from ancestries more readily available in these cohorts, contributing to systemic health disparities. Additionally, datasets often favor the most common cancer subtypes, which in breast cancer inversely correlates with tumor aggressiveness and prognosis. We propose the generation of breast cancer transcriptomic profiles using data from The Cancer Genome Atlas and a denoising diffusion probabilistic model. We will compare our approach to other generative models such as generative adversarial networks. This method enables the creation of synthetic datasets that follow the distributions of real biological data, improving representation of rare molecular subtypes. Additionally, synthetic profiles can provide a practical way to preserve patient privacy while enabling the training of robust genomic models. This approach could help develop more generalizable and privacy-preserving AI tools for precision oncology.

## A-050: Mechanisms driving advanced atherosclerosis in transgenicLPA mice: an scRNA-seq analysis

**Authors List:**
**Sunny Pang**, Department of Biochemistry, University of Western Ontario, Canada
Julia St John, Robart Research Institute; Department of Biochemistry, University of Western Ontario, Canada
Michael Boffa, Robart Research Institute; Department of Biochemistry, University of Western Ontario, Canada
Marlys Koschinsky, Robart Research Institute; Department of Biochemistry, University of Western Ontario, Canada
Tallulah Andrews, Department of Biochemistry; Department of Computer Science, University of Western Ontario, Canada

**Presentation Overview:**
Atherosclerosis is a leading cause of cardiovascular mortality through plaque-driven heart attacks and strokes. Lipoprotein(a) [Lp(a)] is the biggest genetic risk factor, yet the mechanisms by which it promotes lesion progression remain poorly understood. To better understand Lp(a)'s role, we analyzed single-cell RNA sequencing (scRNA-seq) data from whole aortas of human LPA [apo(a)] transgenic mice, with Ldlr disruption and fed 24 weeks of high-fat, high-cholesterol diet. Across all samples, we identified 15 distinct vascular and immune cell types and uncovered sex-specific transcription driven by LPA. Notably, in female LPA mice, resident macrophages showed the highest number of differentially expressed genes, which included upregulation of myofibroblast-related pathways including actin–myosin organization and collagen breakdown alongside downregulation of immune signalling pathways, consistent with a macrophage-to-myofibroblast transition (MMT). RNA velocity analysis confirmed MMT and with key genes such as Acta2, Myl9, and Cald1 increasing further along the trajectory. A similar, though less pronounced, pattern was observed in male LPA mice. Overall, these cell-fate alterations support the enhanced atherosclerosis and more vulnerable plaque features observed in LPA mice relative to controls. Future work could identify therapies to prevent or revert these changes to reduce the burden of atherosclerosis in Lp(a) high individuals.

## A-052: DOTS?RNA: A FAIR?Aligned Infrastructure for mRNATherapeutic Sequence Management and Design

**Authors List:**
**Eric Audemard**, D2R-HeDS, McGill University, Canada
Senthilkumar Kailasam, D2R-HeDS, McGill University, Canada
Guillaume Bourque, D2R-HeDS, McGill University; Department of Human Genetics,McGill University, Canada
Mathieu Bourgey, D2R-HeDS, McGill University, Canada
Sebastian Ballesteros Ramirez, D2R-HeDS, McGill University, Canada
Maryam Youssef, D2R-HeDS, McGill University, Canada

**Presentation Overview:**
The rapid expansion of mRNA?based therapeutics highlights the need for robust, standardized infrastructure to support the Design and Optimization of Therapeutic Sequences (DOTS) across collaborative research environments. We present DOTS?RNA, an open?source, Django?based web platform providing a structured and extensible data foundation for mRNA sequence management, construct design, and computational optimization workflows. DOTS?RNA implements a structured three?tier data model—sequences, constructs, and projects—enabling researchers to manage biological sequences (FASTA/GenBank), assemble complex mRNA constructs (linear mRNA, CAR?T, saRNA, circRNA), and execute containerized optimization pipelines. Global deduplication using MD5 hashing ensures data integrity, while project?scoped metadata supports contextual reuse of shared sequences. The platform provides stable identifiers, rich and automatically computed annotations (e.g., GC content, Kozak motifs, RNA secondary structure), granular access control, and programmatic access through a REST API. Asynchronous execution with Celery and cloud?based infrastructure supports reproducible and scalable analysis workflows. By structuring the accumulation of sequences, annotations, constructs, and optimization outputs, DOTS?RNA lays the groundwork for data? and AI?driven mRNA design, positioning it as foundational infrastructure for reproducible therapeutic sequence research.

**A-054: Niche-Aware Deconvolution of Cell-Type Fractions and Gene Expression from Bulk RNA-seq Data.**

**Authors List:**
**Luis Germán Castañón-Cortés**, Tecnológico de Monterrey, Mexico
Claudia Rangel-Escareño, Tecnológico de Monterrey, Mexico

**Presentation Overview:**
Single-cell RNA-seq has been transformational to transcriptomics; however, this technique remains costly and high-quality bulk RNA-seq data is abundant. Deconvolution methods aim to bring these bulk datasets closer to the single-cell era by computationally extracting cell-type level information from them. In the current landscape, deconvolution methods assume that genes are expressed independently across cell types; however, gene expression is a dynamic function of the local cellular microenvironment, or niche. We found that the errors in cell-type-specific expression estimates are systematically correlated with niche signaling pressure. To address this, we developed a niche-aware deconvolution framework based on BayesPrism. Our method integrates a NicheNet-informed prior to the BayesPrism joint posterior Gibbs sampling, using ligand-target expression correlations from the single-cell reference to define the directionality of niche-driven regulation. Preliminary results using bulk data with matched single-nucleus RNA-seq as ground truth show that our method can reduce BayesPrism's absolute estimation error and recover biological trends across patients, particularly for reactive and rare cell states such as microglia. This study highlights the potential of leveraging intercellular interactions to enhance deconvolution accuracy and uncover hidden biological insights from bulk transcriptomic datasets.

## A-056: Spatial Transcriptomics Reveals Disrupted IntercellularCommunication and Reduced Neuroplasticity Signaling in aRodent Model of Postpartum Depression

**Authors List:**
**Maria Eleni Fafouti**, University of Toronto, Canada
Shreejoy Tripathy, Department of Psychiatry, University of Toronto / Krembil Centre for Neuroinformatics (KCNI), Canada
Shahrzad Ghazisaeidi, Krembil Centre for Neuroinformatics (KCNI), Canada
Hugo Hudson, University of Toronto, Canada
Bharti Kukreja, Boston Chrildren's Hosptial, United States
Liisa Galea, Centre for Addiction and Mental Health (CAMH), Canada
Romina Garcia De Leon, University of Toronto, Canada
Brian Kalish, Boston Chrildren's Hosptial, United States

**Presentation Overview:**
Pregnancy and the postpartum period involve remarkable neuroplasticity but also heightened vulnerability to postpartum depression (PPD). We employed two spatial transcriptomics platforms: Slide-seq and Slide-tags to chart gene expression and cellular remodelling in a corticosterone-induced rodent model of PPD, focusing on limbic regions like the hippocampus. Our results reveal significant shifts in cell-type proportions and disrupted intercellular communication in corticosterone-treated rats. Specifically, we observed a general decrease in signaling between the CA3 region and other hippocampal areas, as well as non-neuronal cells. Transcriptionally, we identified a dual-response mechanism: stress-resilience markers like BDNF were upregulated, yet pathways essential for structural plasticity - including Netrin, Slit, and Pleiotrophin - were significantly downregulated. By integrating these cutting-edge technologies, our study offers a uniquely detailed perspective on the spatial organization of gene expression in the maternal postpartum brain. These findings pinpoint the CA3 microenvironment as a focal point of PPD-related dysfunction and highlight targets to restore peripartum neuroplasticity.

## A-058: Analysis of short-term mutational pressures within cancerin the context of long-term mammalian evolution

**Authors List:**
**Yash Patel**, Northeastern University, Canada
Paulo Miguel, Northeastern University, Canada

**Presentation Overview:**
With advances in genome sequencing technologies, genes associated with cancer initiation and progression have been identified and leveraged to predict susceptibility and prognosis. However, little is known about how genes mutate during tumorigenesis. Viewing cancer as a process of somatic evolution, we hypothesized that patterns from evolutionary genomics could clarify mutation dynamics. We developed a two-part integrated computational framework. The upstream module processed multi-species genome alignments in MAF or gene-level FASTA format using FRESCo, applying codon-based likelihood ratio tests with Bonferroni correction to detect significant rate deviations. It identified regions with elevated synonymous substitution rates, termed Synonymously Accelerated Elements (SAEs), and slowly evolving Synonymously Constrained Elements (SCEs), which served as long-term evolutionary references. The downstream module analyzed cancer mutation data from 18,497 samples, encompassing 3,144,235 single nucleotide polymorphisms, to identify genes with significantly altered mutation rates within SAEs and SCEs. We found 1,031 genes showing differential patterns in SAEs and 941 in SCEs, including 585 shared genes. These findings support a proposed somatic evolutionary model of cancer and motivate further experimental validation. Together, this framework integrates comparative genomics and tumor mutational landscapes to generate testable hypotheses about selective pressures shaping cancer genomes across diverse tissues and evolutionary timescales globally observed.

## A-060: Parn Knockout Induces miRNA and Small RNA Changes in a Zebrafish Model of Dyskeratosis Congenita

**Authors List:**
**Ben Patrick**, University of Ottawa, Canada
Sarada Ketharnathan, CHEO Research Institute | Departments of Pediatrics & Cellular and Molecular Medicine, University of Ottawa, Canada
Arvind Mer, University of Ottawa, Canada
Jason Berman, CHEO Research Institute | Departments of Pediatrics & Cellular and Molecular Medicine, University of Ottawa, Canada

**Presentation Overview:**
Poly(A)-specific ribonuclease (PARN) is a 3? exonuclease required for the maturation and stability of multiple small RNA biotypes. Loss-of-function mutations in parn are implicated in dyskeratosis congenita (DC), a type of inherited bone marrow failure characterized by reduced hematopoiesis and telomere shortening, yet its effect on small RNA processing remains poorly defined. We performed small RNA sequencing in kidney marrow (analogous to human bone marrow) tissue isolated from 12-month-old CRISPR–Cas9 parn-mutant zebrafish, to profile differential expression and non-templated 3’ tailing events. Parn loss predominantly affected micro RNAs (miRNAs), particularly regulators of cell differentiation and hematopoiesis, and was associated with enrichment of pathways reflecting transcriptional restraint and post-transcriptional regulation, consistent with miRNA-driven remodeling of gene expression programs. To quantify tailing events, we developed a computational workflow detecting non-templated 3? nucleotide additions by strand-specific parsing of soft-clipped reads from BAM files. parn mutant marrows exhibited longer miRNA 3? tails (?1nt), driven by increased polyadenylation and polyuridylation, while snoRNAs and scaRNAs were largely unaffected. Non-templated tailing patterns were miRNA-specific, revealing heterogeneous processing. Our results demonstrate biotype-specific disruption of miRNA processing in a zebrafish model of DC and provide a reproducible framework for studying RNA processing defects associated with exonuclease dysfunction.

## A-062: Deep Learning Reveals Divergent Stem Cell Populations Linked to Drug Resistance in Leukemia

**Authors List:**
**Nicolas Ho**, University of Ottawa, Canada
Héléna Boutzen, Princess Margaret Cancer Centre, Canada
John Dick, Princess Margaret Cancer Centre, Canada
Arvind Mer, University of Ottawa, Canada

**Presentation Overview:**
Acute myeloid leukemia (AML) relapse is driven by leukemia stem cells (LSCs) that persist through induction chemotherapy, yet the mechanisms enabling their survival remain incompletely defined. Prior models have invoked cell-intrinsic quiescence, metabolic adaptation, or drug efflux, but clinical trials targeting each of these axes have largely failed to improve outcomes. We hypothesized that unresolved heterogeneity within the hematopoietic stem cell (HSC) compartment underlies differential chemoresistance. We assembled a harmonized single-cell RNA sequencing resource of 523,645 cells across eight independent cohorts spanning primary patient bone marrow aspirates, patient-derived xenografts, and a patient-derived cell line. An integrative pipeline combining reference-based annotation, deep generative modeling via single-cell variational inference, and k-means clustering resolved 9,394 high-confidence HSCs into two transcriptionally distinct subpopulations, i.e.: translationally active HSCs (T-HSCs), enriched for ribosomal biogenesis, protein synthesis, and stress-response programs (NR4A1/2/3, FTH1, CXCR4 etc.), and anchored HSCs (A-HSCs), enriched for integrin- and adhesion-mediated niche engagement (ITGA4, ITGA6, CD99, CALCRL, FLT3 etc.). Cell cycle analysis using both Tirosh-based scoring and the ccAFv2 classifier showed near-identical phase distributions between the two subtypes, indicating that A-HSCs are not distinguished from T-HSCs by quiescence. We derived concise cluster-specific gene signatures and evaluated prognostic performance across four independent bulk AML cohorts (BeatAML, Leucegene, TCGA-LAML, UHN; n = 772). A-HSC signature enrichment predicted significantly worse prognosis (meta-analysis $p = 0.00022$) and remained an independent predictor in multivariate Cox models adjusting for NPM1, FLT3-ITD, TP53, DNMT3A, karyotype risk, age, and sex. In bivariate models, the A-HSC score outperformed or complemented the LSC17 classifier. Paradoxically, the A-HSC signature correlated with greater ex vivo sensitivity to most tested compounds in BeatAML — including cytarabine and multiple FLT3/kinase inhibitors — and A-HSC cells showed no deficit in cytarabine metabolism or cytotoxicity response, supporting niche-mediated rather than cell-intrinsic resistance. Cell-cell communication analysis across three cohorts identified niche signals such as CD99, THBS, and CALCRL to preferentially target A-HSCs, but undetectable toward T-HSCs. Visium spatial transcriptomics of an AML trephine biopsy localized A-HSC-dominant spots directly at the endosteal surface, while T-HSC spots were significantly displaced from both bone and adipocyte compartments. A-HSC burden further increased at relapse and was elevated at diagnosis in patients who failed to achieve complete remission. Together, these results identify a clinically relevant niche-anchored stem-like population whose chemoresistance reflects microenvironmental protection rather than intrinsic dormancy or drug resistance, and nominate the A-HSC niche-engagement programs as actionable therapeutic vulnerabilities.

## A-064: Achieving T2T Resolution in the NCLCN-261 iPSC Line via Multi-Platform Integration

**Authors List:**
**Robert Eveleigh**, McGill University / Canadian Centre for Computational Genomics (C3G), Canada
Mitinori Saitou, Kyoto University / Advanced Study of Human Biology (WPI-ASHBi), Japan
Yukihiro Yabuta, Kyoto University / Advanced Study of Human Biology (WPI-ASHBi), Japan
Guillaume Bourque, McGill University / C3G and Kyoto University / ASHBi, Canada

**Presentation Overview:**
High-quality, telomere-to-telomere (T2T), haplotype-resolved genome assemblies form the foundation of modern pangenome construction, enabling mitigation of reference bias and accurate resolution of structurally complex regions such as centromeres and segmental duplications. These assemblies provide a robust ground truth that supports a shift from isolated variant detection toward comprehensive genome inference across diverse human populations. We demonstrate this progress through the collaborative assembly of the NCLCN-261 induced pluripotent stem cell (iPSC) line. Through iterative refinements using the latest assembly algorithms integrating 81× PacBio Revio, 104× ONT ultra-long (44× >100 kb), and 44× Duplex Pore-C data, we achieved incremental improvements in contiguity and completeness. The current assembly comprises 50 scaffolds per haplotype, with mean scaffold lengths of 121.1 Mbp (hap1) and 120.7 Mbp (hap2), reflecting reduced fragmentation and improved gap resolution. These refinements enabled the identification of 24 T2T scaffolds, 15 T2T contigs, generating 7 fully resolved T2T chromosomes. Gene completeness analyses further indicate improved recovery of previously missing genes relative to earlier assembly iterations. Together, these advances underscore the importance of highly contiguous, chromosome-scale assemblies for extending genomic medicine into previously inaccessible regions of the human genome and for advancing the development of a more representative and accurate global human pangenome.

## A-066: Reactome: A Trusted, Open Knowledgebase for IntegrativePathway Bioinformatics

**Authors List:**
Nancy Li, OICR, Canada
**Reactome Consortium**, OICR, Canada

**Presentation Overview:**
Reactome: A Trusted, Open Knowledgebase for Integrative Pathway Bioinformatics Reactome is a freely accessible, open-source, expert-curated knowledgebase of human biological pathways that represents molecular processes as structured, computable reaction networks. It functions as both a digital textbook of molecular mechanisms and a computational platform for interpreting high-throughput genomics, transcriptomics, proteomics, and multi-omics datasets. Pathways spanning signaling, metabolism, gene regulation, immune function, and disease are manually curated by domain experts, peer-reviewed, and supported by primary literature, ensuring high-quality, biologically accurate annotation. Recent developments have expanded pathway coverage, enhanced disease and genomic variant annotation, and strengthened integration of drug–target interactions. Updated visualization tools, including interactive high-level diagrams and global pathway overviews, enable intuitive exploration of complex systems. Improved analysis workflows support pathway enrichment, comparative cross-species inference, and programmatic access via APIs, facilitating scalable and reproducible computational research. Reactome adheres to FAIR data principles and is recognized as both an ELIXIR Global Core Biodata Resource and a CoreTrustSeal-certified repository, reflecting its commitment to long-term sustainability, transparency, and rigorous data stewardship. As an open community resource, Reactome actively invites collaboration from the Canadian bioinformatics community to contribute expertise, expand pathway coverage, and advance shared infrastructure for systems biology and precision medicine.

## A-068: Characterizing and Mitigating Protocol-Dependent Gene Expression Bias in 3? and 5? Single-Cell RNA Sequencing

**Authors List:**
**Valeryia Shydlouskaya**, Western University, Canada
S.M. Mansour Haeryfar, Western University, Canada
Tallulah S. Andrews, Western University, Canada

**Presentation Overview:**
Single-cell RNA sequencing (scRNA-seq) enables large-scale characterization of cellular heterogeneity, yet integrating datasets generated using different library preparation protocols remains challenging. Comparisons between 10X Genomics 3? and 5? chemistries are complicated by protocol-specific technical biases arising from differences in transcript end capture and amplification. Although normalization and batch correction are common scRNA-seq preprocessing steps, it remains unclear which correction is most appropriate, or whether it is necessary, for reliable cross-protocol comparison. Here, we characterize protocol-related expression differences using 35 matched donors across six tissues profiled with both 3? and 5? scRNA-seq. We find that gene expression discrepancies are not pervasive across the whole transcriptome but are driven by a small, reproducible subset of protocol-biased genes. Excluding these genes improves cross-protocol concordance, indicating that most genes are comparable without aggressive correction. In benchmarking commonly used normalization approaches, we show that while several methods improve statistical alignment when cell populations are well matched, they can distort gene-level signals and inflate differential expression in biologically realistic settings with incomplete cell-type overlap. Taken together, our results demonstrate that protocol bias between 3? and 5? scRNA-seq is limited and that targeted handling of biased genes offers an alternative to broad normalization or batch correction strategies.

## A-070: A New Model for Analyzing Mutation Accumulation in Viruses

**Authors List:**
**Jake Daigle**, Western University, Canada
Art Poon, Western University, Canada

**Presentation Overview:**
Populations that undergo repeated and severe bottlenecks, such as viruses transmitted between hosts, can accumulate mutations irrespective of their effects on fitness. This stochastic erosion of fitness under mutation accumulation can be a significant driver of extinction. Our theoretical understanding of this process is largely restricted to models with the unrealistic assumption that the effect of any mutation is independent of its genomic context and environment. We describe a new stochastic model in which the probability P that the next mutation is deleterious (reduces fitness) is a function of current fitness. Our model generalizes previous models in which a mutation's effect on fitness depends on the current phenotype (Fisher's geometric model) or genotype (the Rough Mount Fuji model), while being more feasible to parameterize. We used Rust to implement an agent-based simulation model of a population undergoing mutation and bottlenecks to explore the relationships between the model parameters (e.g. the functional form of the relationship between fitness and P) and the expected time to extinction. Our results indicate that the relationship between fitness and P can be substantial and non-linear. These models provide a novel analytical framework for understanding viral population dynamics.

## A-072: From Chromosomal Loss to EMT: Exploring the Role of 4p inTNBC Dedifferentiation

**Authors List:**
**Zahra Tavakol Hamedani**, McGill University, Canada
Hamed S. Najafabadi, McGill University, Canada
Jean Monlong, INSERM U1220 IRSD, France
Elena Kuzmin, McGill University/ Concordia University, Canada

**Presentation Overview:**
Large chromosomal deletions are common in cancer, but their functional impact on tumour cell identity is often unclear. Basal breast cancer is an aggressive subtype in which chromosome 4p (chr4p) loss is one of the most frequent genomic events and is linked to poor clinical outcome. However, how this deletion shapes transcriptional programs in primary tumours has not been systematically defined. Using single-cell transcriptomic profiles from four treatment-naïve triple-negative breast cancer datasets from the Curated Cancer Cell Atlas, we show that epithelial tumour cells with chr4p loss occupy a distinct gene expression state consistent with increased plasticity. Using curated epithelial markers, we selected malignant epithelial populations and used normal breast epithelial cells as a reference. We then inferred copy-number states at single-cell resolution and identified tumour cell communities with chr4p loss and copy-neutral profiles. Comparison of these populations showed a transcriptional program associated with chr4p loss, including enrichment of epithelial–mesenchymal transition, inflammatory signaling, and developmental pathways. These results suggest that chr4p loss is not only a recurrent genomic marker, but is associated with coordinated changes in tumour cell differentiation programs in primary disease. Broadly, our study provides a framework for linking recurrent arm-level copy-number losses to tumour cell-state transitions.

## A-074: Where Data Meets Meaning: Ontology Integration forPathogen-Genomics Contextual Data

**Authors List:**
**Rhiannon Cameron**, Simon Fraser University, Canada
Damion Dooley, Simon Fraser University, Canada
Emma Griffiths, Simon Fraser University, Canada
William Hsiao, Simon Fraser University, Canada

**Presentation Overview:**
The SARS-CoV-2 pandemic highlighted to the public health community the need to collect and compare viral genome contextual data. However, contextual data across the globe often varies in structure, value, and format, making meaningful comparisons and analyses difficult. To resolve this issue and improve pathogen surveillance, our team has created open-source, ontologized contextual data specifications that harmonize data from disparate sources. Data harmonization reconciles differences between data streams while data specifications are guidelines that formalize how data should be structured, formatted, and described. Working with Open Biological and Biomedical Ontologies Foundry domain ontologies, specification fields and picklist values are being integrated into the open-source OBOF Genomic Epidemiology Ontology to provide an interoperable and controlled vocabulary that is responsive to dynamic stakeholders needs. An ontology is a machine-actionable representation of a subject that encapsulates its entities, properties, and the relationships between them; they offload modeling to computers, freeing up researchers to focus on analyses. This methodology has been used to ontologize 100s of fields and 1000s of picklist values for national and international data specifications. This work transforms pathogen-genomic specifications from basic data structures and vocabularies into ontology constructs that support computational reasoning; cross-jurisdiction comparison; and scalable, community-adaptable data infrastructure.

## A-076: Regional Canadian Bioinformatics Hub (CBH) Activities in British Columbia & Ontario

**Authors List:**
**Tian Rabbani**, Simon Fraser University (SFU) | Canadian Bioinformatics Hub (CBH), Canada
Neha Ratti, University Health Network (UHN) | Canadian Bioinformatics Hub (CBH), Canada
Karen Cranston, University Health Network (UHN), Canada
Michael Brudno, University Health Network (UHN), Canada
William Hsiao, Simon Fraser University (SFU), Canada

**Presentation Overview:**
The Canadian Bioinformatics Hub (CBH) leads training and community initiatives in Bioinformatics, Computational Biology, Life and Health Sciences, and Data Science (BCBDS) across six regions in Canada. Given the country's decentralized structure and the challenges of information silos, strong regional coordination is essential to advance bioinformatics career development and growth. In 2025, the CBH offered four workshops across three regions, with plans for more workshops across all regions in 2026. In British Columbia (BC), workshops on R and machine learning have been instrumental in connecting various BCBDS communities across Vancouver, Burnaby, and Surrey. In 2026, BC and Ontario will host workshops aligned with regional priorities and local research needs across BCBDS communities. Offerings will range from foundational courses to advanced topics including bulk and single-cell RNA sequencing analysis, and cancer genomics. Capacity building will be strengthened through seminars hosted by local Bioinformatics Users Groups, VanBUG and TorBUG, as well as other partners creating valuable mentorship and networking opportunities for trainees. To further enhance connectivity in Ontario and BC, the Network of Champions initiative will continue bringing together faculty, students, and research groups across institutions to act as local advocates to foster collaboration, knowledge exchange, and sustained engagement across these provinces.

## A-078: Mechanistic Characterization of Chromatin Variants via Physics-Informed Geometric Deep Learning

**Authors List:**
**Yiru Sheng**, Department of Biomedical and Molecular Sciences, Queen's University, Kingston, Canada, Canada
Sunghwan Moon, Department of Pathology and Molecular Medicine, Queen's University, Kingston, Canada, Canada
Anna Panchenko, Department of Biomedical and Molecular Sciences, Queen's University; Ontario Institute of Cancer Research, Canada

**Presentation Overview:**
Quantifying the mechanistic impact of mutations on chromatin dynamics remains a critical challenge in chromatin biology. Most widely used variant effect predictors effectively leverage sequence conservation and sequence-context signals. Even state?of?the?art predictors that leverage structural context at the level of single proteins, these remain limited and do not capture the broad architecture or interaction networks of large, multi?complex macromolecular assemblies. We present a physics-informed geometric deep learning framework to define the mutational "zone of influence"—the structural neighborhood perturbed by a mutation. We model the chromatin interactome as a graph derived from atomistic chromatin-related structures. To capture differential interaction contexts, we employ a Siamese Graph Neural Network trained via contrastive learning, optimizing embeddings to distinguish mutant-induced structural perturbations from wild-type baselines. Furthermore, we implement a differentiable heat diffusion module where edge weights are initialized from interaction energies and fine-tuned to map mutation permutation propagation pathways. Preliminary results indicate that pathogenic residues are non-randomly distributed, preferentially localizing to network hubs and bottlenecks that control long range signaling. This method systematically identifies mechanistically prioritized variants and communication pathways connecting mutations to distal functional interfaces, offering a robust tool for structural systems biology and therapeutic targeting.

## A-080: FoodOn 2.0: A Scalable Ontology for the InteroperableRepresentation of Food Data Across Domains

**Authors List:**
Anoosha Sehar, The Centre for Infectious Disease Genomics and One Health,Simon Fraser University, Canada
Damion Dooley, The Centre for Infectious Disease Genomics and One Health,Simon Fraser University, Canada
Kai Blumberg, USDA, United States
**William Hsiao**, The Centre for Infectious Disease Genomics and One Health,Simon Fraser University, Canada
Emma Griffiths, The Centre for Infectious Disease Genomics and One Health,Simon Fraser University, Canada
Rhiannon Cameron, The Centre for Infectious Disease Genomics and One Health,Simon Fraser University, Canada
Charlie Barclay, The Centre for Infectious Disease Genomics and One Health,Simon Fraser University, Canada

**Presentation Overview:**
Food data spans production, processing, consumption, health, and research, yet inconsistent terminology and data structures limit integration and reuse. Research, regulatory, and public-sector communities, therefore, require shared semantic frameworks that allow foods and food-related data to be described consistently across heterogeneous sources. FoodOn is an open-source ontology developed to support this need by providing a standardized representation of foods, food materials, and related concepts. Here we describe FoodOn 2.0, a conceptual and structural update designed to improve scalability, consistency, and interoperability across evolving food datasets. FoodOn 2.0 introduces a refined upper-level organization that distinguishes organisms, organism-derived materials, food materials, and food products, enabling clearer representation of food states, roles, and intended uses. The resource covers foods derived from animals, plants, fungi, and seafood, while aligning with biological taxonomy and anatomy. To support sustainable ontology growth, FoodOn 2.0 adopts structured and repeatable modeling patterns developed through community collaboration. These patterns facilitate integration with external knowledge sources, including nutrient databases and dietary ontologies. By improving consistency and reusability, FoodOn 2.0 helps make food data more findable and comparable. Overall, FoodOn 2.0 demonstrates how scalable knowledge infrastructure supports cross-domain collaboration and applied research in public health and nutrition.

## A-082: AN ANCHOR-BASED FRAMEWORK FOR CONFORMAL PREDICTION IN NOISYDATA

**Authors List:**
**Morgan He**, University of Western Ontario, Canada
Grace Yi, University of Western Ontario, Canada
Pingzhao Hu, University of Western Ontario, Canada

**Presentation Overview:**
Conformal prediction provides a flexible framework for quantifying predictive uncertainty and has attracted growing interest in machine learning. However, most existing methods are developed for cleanly labeled data and can don't perform well in the presence of label noise. In this work, we consider when ground-truth labels are unobserved and only crowdsourced noisy annotations are available. We propose an anchor-based conformal prediction framework for uncertainty quantification under noisy labeling. The method identifies anchor points by selecting samples that exhibit strong agreement across annotators and uses these anchors to train a base predictor for conformal calibration. This enables the construction of predictive sets that achieve a desired coverage level while mitigating the impact of annotation noise. In addition, we provide a theoretical analysis of anchor-point identification and establish conditions under which the proposed approach is valid, highlighting assumptions that have been largely overlooked in prior work. We demonstrate the effectiveness of our method on single-cell RNA sequencing datasets, where it achieves reliable coverage with more informative prediction sets compared to standard conformal baselines.

## A-084: Decoding the VHL-pRb Axis in Clear Cell Renal Cell Carcinoma

**Authors List:**
**Max Jarry**, University of Ottawa, Department of Biochemistry, Microbiology and Immunology, Canada
Bahareh Hekmatnejad, University of Ottawa, Department of Cellular and Molecular Medicine, Canada
Ryan Russell, University of Ottawa, Department of Cellular and Molecular Medicine, Canada
Arvind Mer, University of Ottawa, Department of Biochemistry, Microbiology and Immunology, Canada

**Presentation Overview:**
Clear-Cell Renal Cell Carcinoma (ccRCC) is the most common subtype (~80%) of RCC. In ccRCC, biallelic Von Hippel-Lindau (VHL) gene defects are a hallmark characteristic. The VHL tumour suppressor is a substrate recognition subunit found in E3 ubiquitin ligase complexes involved in proteasome-mediated degradation. Interestingly, recent research has uncovered novel VHL-related pathways where the dysregulation of the retinoblastoma protein (pRb) due to VHL loss also promotes ccRCC through potential interactions with a chromatin remodeler (CR). To define the contribution of the VHL-pRb axis to ccRCC progression, we generated VHL and pRb knockout ccRCC cell lines and performed RNA sequencing. Comparative transcriptomic analysis revealed enrichment of pathways associated with the candidate CR, chromatin organization, and broader epigenetic regulation. To further characterize these mechanisms, we are implementing multi-omics profiling, including ATAC-seq and CUT&RUN, to map changes in chromatin accessibility and regulatory factor occupancy driven by VHL and pRb loss. Integrating these datasets will clarify how VHL-pRb dependent chromatin remodeling promotes ccRCC and may identify actionable vulnerabilities to help overcome therapeutic resistance and improve outcomes in metastatic disease.

## A-086: Variation in an immune gene family associated with aviancholera survival in an Arctic duck species

**Authors List:**
**Eva Innocente**, Queen's University Department of Biology, Canada
Shayla Kroeze, Queen's University Department of Biology, Canada
Grant Gilchrist, National Wildlife Research Centre, Environment and Climate Change Canada, Canada
Oliver Love, The University of Windsor, Canada
Mark Forbes, Department of Biology, Carleton University, Canada
Vicki Friesen, Queen's University Department of Biology, Canada

**Presentation Overview:**
Infectious diseases are an increasing threat to wildlife and can lead to population decline or risk of extinction. Genomic methods can elucidate the contributions of immune gene variation to disease resistance, but are understudied in wild populations. Between 2006–2008, an avian cholera (Pasteurella multocida) outbreak at Mitivik Island, Nunavut, killed 30% of breeding female Common Eiders (Somateria mollissima). We sequenced whole genomes from eiders sampled before and after the cholera epizootic and employed bioinformatic methods to characterise inter-individual variation in an immune gene family (toll-like receptors; TLRs). Across all single nucleotide polymorphisms (SNPs), greater heterozygosity of TLR3 was significantly associated with cholera survival. Exclusion of synonymous SNPs revealed an association between greater TLR4 heterozygosity and survival that approached significance. Several TLR SNPs were significantly associated with survival, and other SNPs showed evidence of selection. Frequency of one TLR21 amino acid was significantly higher in surviving eiders. Results suggest a genetic basis for avian cholera resistance among some Common Eiders. These results could inform future disease management and conservation efforts for eiders and other Arctic species facing exposure to novel pathogens and highlights the use of bioinformatics for understanding the possible impacts of disease among species in the wild.

## A-088: Tissue-specific genetic and epigenetic characterization of maternal plasma cell-free DNA

**Authors List:**
**Natalie Yuen**, McMaster University, Canada
Melanie Lemaire, McMaster University, Canada
Samantha Wilson, McMaster University, Canada

**Presentation Overview:**
During pregnancy, cell-free DNA (cfDNA) in maternal blood is comprised of a mixture of maternal and placental fragments. Most analyses do not differentiate these fragments, resulting in masked tissue-specific signals. We aim to classify the tissue-of-origin of cfDNA fragments in maternal plasma using genetic marks and characterize tissue-specific epigenetic patterns throughout pregnancy. Maternal blood (n = 24) was collected between 11-39 weeks of pregnancy. Plasma cfDNA was sequenced using biomodal duet evoC, which measures genetic sequence, DNA methylation and DNA hydroxymethylation. Maternal single nucleotide polymorphism (SNP) genotypes were obtained through buffy coat DNA variant calling. Placenta-specific SNP alleles, identified by scanning cfDNA for non-maternal alleles, were used to distinguish placental cfDNA fragments. On average, 0.51% ± 0.21% of cfDNA fragments were classified as placental. Placental cfDNA fractions were positively correlated with gestational age ($r = 0.49$, $p = 0.02$) and RASSF1 promoter DNA methylation ($r = 0.47$, $p = 0.02$), a known placental epigenetic mark. We demonstrate that placenta-specific alleles are detectable in maternal plasma cfDNA and can be used to identify placental cfDNA fragments. Next, we will incorporate maternal-specific SNP alleles, epigenetic marks, and SNP imputation to increase the genomic coverage of tissue-specific cfDNA fragments for epigenetic characterization.

## A-090: Toronto Bioinformatics Group (TorBUG): From Local Hub to National Node

**Authors List:**
**Robin Haw**, Ontario Institute for Cancer Research | TorBUG, Canada
Torbug Team, TorBUG, Canada

**Presentation Overview:**
Bioinformatics User Groups (BUGs) are key to community building in Canadian bioinformatics, uniting trainees, researchers, and professionals passionate about computational biology and data science. Established in 2012, the Toronto Bioinformatics Users Group (TorBUG) fosters knowledge exchange between academia and industry. Strategically situated at the University of Toronto, TorBUG leverages a thriving research ecosystem of institutions, departments, hospitals, industry, and pharma. This proximity facilitates opportunities for seamless knowledge transfer, collaboration, and networking. Originally in-person, TorBUG pivoted to a fully online model during the pandemic and later evolved into the current hybrid format. This approach ensures accessibility, allowing participants to engage from across Canada and the World. Traditionally, we pair guest speakers with trainee talks at a monthly seminar. This helps graduate students gain visibility while keeping the community up to date on their research. We introduced Trainee Speaker Rounds to increase cross-departmental visibility and Career Fireside Chats, where professionals share candid career narratives. In alignment with our commitment to inclusivity, all events are free and open to the research community. TorBUG strengthens the national Canadian Bioinformatics Hub (CBH) by collaborating with established nodes such as VanBUG, MonBUG, and BioNet, as well as newly launched groups such as yegBUG and Batl.

## A-092: Preparing data for a long journey: early curation helps researchers generate more reusable, high quality datasets

**Authors List:**
**Ruobin Liu**, University of Saskatchewan, Canada
Lacey-Anne Sanderson, University of Saskatchewan, Canada
Carolyn T. Caron, University of Saskatchewan, Canada
Reynold Tan, University of Saskatchewan, Canada
Vidusha Wilpita, University of Saskatchewan, Canada
Laura Jardine, University of Saskatchewan, Canada
Kirstin Bett, University of Saskatchewan, Canada

**Presentation Overview:**
Data is a fundamental element in research of any kind. While large data repositories exist for genomics data, how can researchers better support all stages of a research experiment, from generation, backup, publication, and reuse? Some challenges we have encountered include inconsistencies in data collection method, ambiguity in links between datasets, and miscommunication on custom file formats. Our work takes a wide range of data inputs including, but not limited to, raw agricultural field trials, biochemical assays, and high throughput imaging outputs. These data are backed up during active collection and then curated both automatically via validation tools built into database importers, and manually by a curator in collaboration with the data collectors. This ensures consistency and completeness of the data, including metadata, prior to being stored in a relational database with FAIR principles in mind. Furthermore, having these high-quality datasets stored in a database allows seamless integration with bioinformatic and data visualization tools, generation of dynamic pages describing datasets and facilitating deeper insights into research data through future reuse. The resulting system not only enhances current biological insight but also creates a sustainable foundation for future research, collaboration, and innovation.

## A-094: Future-Proofing Canadian Data Infrastructure ThroughAdaptive System Design: The iMicroSeq Data Portal Example

**Authors List:**
**Justin Richardsson**, Ontario Institute for Cancer Research, Canada
Leonardo Rivera, Ontario Institute for Cancer Research, Canada
Edmund Su, Ontario Institute for Cancer Research, Canada
Bhavik Bhagat, Ontario Institute for Cancer Research, Canada
Jose Hector Galvez, McGill University / Canadian Centre for Computational Genomics, Canada
Baofeng Jia, Simon Fraser University, Canada
Nima Tehrani, Simon Fraser University, Canada
Charlie Barclay, Simon Fraser University, Canada
Erin Gill, Simon Fraser University, Canada
Emma Griffiths, Simon Fraser University, Canada
Guillaume Bourque, McGill University / Canadian Centre for Computational Genomics, Canada
Fiona Brinkman, Simon Fraser University, Canada
Melanie Courtot, Ontario Institute for Cancer Research, Canada

**Presentation Overview:**
When platforms cannot flexibly adapt to evolving data modalities, they become constrained, limiting contributor participation, data discoverability, and scientific potential. Building custom systems for each new requirement or data source compounds these limitations with unsustainable development costs, threatening long-term project viability. The iMicroSeq Data Portal (https://imicroseq-dataportal.ca), developed by a pan-Canadian academic-public health collaboration, with government and community-based input, demonstrates how upfront investment in reusable architecture maximizes long-term returns on data infrastructure. Here we describe how early architectural decisions enabled flexible platform extension to support new data modalities. Building upon the existing VirusSeq Data Portal (>650,000 SARS-CoV-2 viral genomes), we extended the underlying components to incorporate wastewater and other water environmental sequence data from diverse microbial species without modifying core services or rewriting interfaces. This was achieved with Overture (https://overture.bio), a software stack designed to separate configurable domain logic (metadata schemas, validation rules, search interface models) from reusable core services and functionalities (data-submission, validation, storage, exploration). iMicroSeq now provides a unified platform for aggregating and exploring heterogeneous clinical and environmental datasets while maintaining FAIR-aligned data management. Our approach reduced engineering overhead while enabling agile, collaborative surveillance infrastructure that adapts as priorities and sample types evolve.

## A-096: A Systematic Review Mapping on Quantum Bioinformatics

**Authors List:**
**Mehdi Khalaj**, Department of Computer Science, University of Saskatchewan, Saskatoon, SK, Canada
Steven Rayan, Centre for Quantum Topology and Its Applications (quanTA), University of Saskatchewan, Saskatoon, SK, Canada
Lingling Jin, Department of Computer Science, University of Saskatchewan, Saskatoon, SK, Canada

**Presentation Overview:**
Modern bioinformatics faces escalating challenges stemming from both the inherent computational hardness of many fundamental problems and the rapidly growing scale and complexity of biological data, increasingly limiting the effectiveness of classical computational approaches. These challenges demand sophisticated computational biology approaches, including algorithmic advances, to generate candidate hypotheses and mechanistic insights, which can then be experimentally validated through wet-lab studies to establish biological relevance. Quantum computing has emerged as a promising paradigm for addressing these challenges by enabling alternative problem representations and novel search strategies for exploring complex solution spaces. We present a systematic mapping review of quantum bioinformatics, synthesizing studies across ten bioinformatics domains. We categorize quantum and hybrid quantum–classical algorithms, analyze problem formulations and encoding strategies, and examine how current research addresses Noisy Intermediate-Scale Quantum (NISQ) constraints such as noise, limited qubit counts, and scalability. Our analysis highlights dominant methodological patterns, emerging application areas, and recurring limitations. The results and the data analysis indicate that most studies remain proof-of-concept and simulation-based, with scalability and hardware robustness as key barriers. This review provides an integrated perspective on the field and outlines critical directions for advancing quantum bioinformatics toward practical biological impact.

## A-098: Attention-guided deep learning integration of multiomicdata in pancreatic cancer

**Authors List:**
**Fatemeh Behjati**, Western University, Canada
Emilie-Jaune Pons, Western University, Canada
Brian Yan, Victoria Hospital, Canada
Michael Sey, Victoria Hospital, Canada
Jennifer Campbell, Victoria Hospital, Canada
Nadeem Hussain, Victoria Hospital, Canada
Ken Leslie, Victoria Hospital, Canada
Zachary Klassen, University of Western Ontario, Canada
Christopher Pin, University of Western Ontario, Canada
Parisa Shooshtari, University of Western Ontario, Canada

**Presentation Overview:**
Pancreatic ductal adenocarcinoma (PDAC) exhibits extreme heterogeneity and resistance to therapy, driven in part by complex epigenetic regulation within the tumor microenvironment (TME). To investigate how cancer-associated fibroblasts (CAFs) influence the epigenetic state of tumor cells, we generated multi?omic data from patient?derived organoids (PDOs) treated with conditioned media from inflammatory (iCAF) and myofibroblastic (myCAF) subtypes. We developed a dual?channel attention?based variational autoencoder (VAE) to integrate RNA?seq and ATAC?seq profiles and identify regulatory mechanisms linked to CAF signaling. iCAF and myCAF chromatin accessibility data were modeled as separate channels embedded within a shared latent space to capture subtype?specific contributions to PDAC heterogeneity. The model's attention mechanism and regression head provided interpretable feature weights that pinpointed chromatin regions most predictive of transcriptional output. Our framework demonstrated high predictive accuracy, uncovering transcription factors (TFs) and regulatory regions epigenetically active yet transcriptionally silent, revealing potential epigenetic reprogramming. Attention?derived regions revealed distinct chromatin signatures associated with key TFs, whose expression and binding patterns suggest candidate targets for epigenetic therapy. We further developed a web platform for interactive exploration of these computational insights. This interpretable deep learning approach advances mechanistic understanding of PDAC epigenetics and supports data?driven strategies for precision oncology.

## A-100: Integrated multi-omics approach for the characterization ofno specific molecular profile in endometrial carcinoma

**Authors List:**
**Farbod Moghaddam**, Department of Obstetrics and Gynecology, UBC, Canada
Dawn Cochrane, Department of Obstetrics and Gynecology, UBC, Canada
Jessica McAlpine, Department of Obstetrics and Gynecology, UBC, Canada
Lien Hoang, Department of Pathology & Laboratory Medicine, UBC, Canada
Andrew Roth, Department of Pathology & Laboratory Medicine, UBC, Canada
Aline Talhouk, Department of Obstetrics and Gynecology, UBC, Canada

**Presentation Overview:**
Background: The ""No Specific Molecular Profile"" (NSMP) subtype of EC is the most common type (50%). While generally associated with good prognosis, NSMP remains heterogeneous and includes patients with very poor outcomes. We utilized multi-omics to resolve this heterogeneity and stratify patient risk. Methods: We clustered mutation and copy-number data from 599 NSMP patients (MSK) utilizing Similarity Network Fusion. A Random Forest classifier (5-fold CV) trained on discriminatory features was applied to two independent TCGA validation cohorts (n=266; n=136). Survival was assessed via log-rank tests. Results: Three clusters were identified, and an RF model achieved 85.6% accuracy (internal validation) using ten features dominated by mutations in the PI3K/AKT and Wnt signalling pathways (e.g., CTNNB1, KRAS, ARID1A, PIK3CA, PTEN). The RF model significantly stratified Overall Survival in the larger TCGA cohort (p=0.02). Cluster 1 (n=7) exhibited the best outcomes, while Cluster 3 (n=176) showed the poorest. The smaller cohort (n=136) displayed comparable Event-Free Survival trends (Cluster 1 best, Cluster 3 worst) but did not reach significance (p=0.33). Conclusion: We validated three molecularly distinct NSMP subgroups with divergent survival outcomes. Future work will focus on confirming that these groups are independently prognostic beyond clinicopathological features and on fully characterizing their biological significance.

## A-102: CBH activities in the Prairies Provinces

**Authors List:**
**Riya Roy**, University of Manitoba, Canada
Athanasios Zovoilis, University of Manitoba, Canada
Eric Merzetti, University of Manitoba, Canada
Sani-E-Zehra Zaidi, University of Manitoba, Canada
Travis Haight, University of Manitoba, Canada
Liam Mitchell, University of Manitoba, Canada
Elly Wu, University of Manitoba, Canada
Angeliki Pantazi, University of Manitoba, Canada
Mohamed Helmy, University of Saskatchewan, Canada
Nia Hughes, Canadian Bioinformatics Hub, Canada
Larisa M. Soto, Canadian Bioinformatics Hub, Canada
Sydney Morgan, Simon Fraser University, Canada
Alexandra Smith, Canadian Bioinformatics Hub, Canada
Michelle Brazas, Canadian Bioinformatics Hub, Canada

**Presentation Overview:**
The bioinformatics ecosystem in Manitoba and Saskatchewan plays an increasingly important role in strengthening the local economy, with demand for highly trained bioinformaticians. The Canadian Bioinformatics Hub (CBH) addresses these gaps through coordinated regional programming and inclusive capacity-building initiatives. Through its Prairie-focused node, CBH delivers hands-on workshops tailored to regional research needs. In 2025, one workshop was held in Prairies, with four planned for 2026. Geographic dispersion across the prairie provinces limits collaboration and knowledge exchange. CBH mitigates this by leveraging BioNet, a regional bioinformatics and omics network, along with a bi-annual regional conference and a monthly seminar series that connect institutions across provinces. These initiatives foster cross-sector collaboration between academia and industry. It has also partnered with local centres and facilities, the University of Manitoba Statistical Genomics and Bioinformatics Platform and the Cancer Care Manitoba Bioinformatics core, to offer five 3-hour workshops in basic digital skills. In partnership with these cores, it also organized a local bioinformatics symposium. A key pillar of this work is engaging Indigenous partners on issues such as genomic and bioinformatics data sovereignty. Collectively, these initiatives enhance bioinformatics expertise, foster regional collaboration, and build more inclusive, resilient bioinformatics and omics communities across the prairies.

## A-104: A Deep Learning Framework for High-Resolution 12S rDNATaxonomic Classification in Fish Metabarcoding

**Authors List:**
**Sk Injamamul Islam**, Department of Biology, Canadian Rivers Institute, University of Prince Edward Island, Charlottetown, PE, Canada, Canada
Michael R. Van Den Heuvel, Department of Biology, Canadian Rivers Institute, University of Prince Edward Island, Charlottetown, PE, Canada, Canada

**Presentation Overview:**
Environmental nucleic acid metabarcoding has become a powerful approach for biodiversity assessment, species monitoring, and early disease detection. However, accurate taxonomic assignment from next generation sequencing data remains a major challenge, particularly for fish studies targeting 12S rDNA, for which no classifier has been explicitly optimized. In this study, we developed a k-mer based deep learning framework for 12S rDNA taxonomic classification. A curated reference dataset was generated from the MitoFish database using the CRABS pipeline, followed by NCBI taxonomy assignment, in silico PCR, and alignment-based refinement. Sequence composition was encoded using frequency chaos game representation at k-mer sizes of 6 and 8. Four deep learning architectures were evaluated, including Convolutional Neural Network, Deep Belief Network (DBN), CNN combined with Bidirectional Long Short-Term Memory, and CNN combined with Transformer, and benchmarked against BLAST. Using a subset of 1000 sequences, the DBN achieved the highest accuracy, reaching 95% at k=6 and 93% at k=8 at the species level. Although BLAST showed comparable overall accuracy, its performance declined at higher taxonomic ranks. The DBN demonstrated greater stability and taxonomic resolution, establishing a robust framework for high resolution 12S rDNA metabarcoding in fish community monitoring and ecological genomics.

**A-106: Chimeras Unleashed: A TE-Aware Multi-Mapping StrategyReveals DNMT1-Driven Disruption of Transposable Element Transcription**

**Authors List:**
**Carlos-Michel Mourra-Diaz**, CHU Sainte-Justine, Canada
Elizabeth Elder, CHU Sainte-Justine, Canada
Serge McGraw, CHU Sainte-Justine, Canada

**Presentation Overview:**
Transposable elements (TEs) are repetitive DNA sequences that can copy and insert across the genome. They are normally epigenetically silenced, yet can be transiently expressed in early development. When this control fails, TE activation can disrupt gene regulation, generate aberrant transcripts, and compromise genome stability. Here, we built a bioinformatics framework to profile TE transcription and TE–gene chimeric transcripts in mouse embryonic stem cells using RNA-seq from a Tet-off system enabling controlled DNMT1 expression. To address TE-specific challenges (sequence redundancy and multi-mapping reads), the pipeline integrates RNA-seq processing with TE-aware quantification (TEtranscripts) and chimeric transcript detection (ChimeraTE). Contrasts across control, DNMT1 inactivation, and rescue conditions generate catalogs of dysregulated TE families and chimeric transcripts. Candidate loci are prioritized by concordant shifts in TE activity and nearby gene expression and genomic context. Ongoing extensions will integrate DNA methylation and histone modification profiles to link chromatin-state changes with TE transcription and chimera formation, producing multi-omic signatures of DNMT1-dependent TE regulation. Biologically, this work pinpoints TE loci that rewire transcription upon DNMT1 loss and normalize upon rescue, clarifying mechanisms that shape early developmental programs, enabling functional tests of impacts on cell identity, developmental robustness, and disease-relevant transcriptional programs.

## A-108: The Bioinformatics Atlantic Network (BAtl): ConnectingTrainees and Professionals in Atlantic Canada

**Authors List:**
**Benjamin Fisher**, Canadian Bioinformatics Hub; Faculty of Computer Science,Dalhousie University, Canada
Robert Beiko, Faculty of Computer Science, Dalhousie University, HalifaxNS, Canada
Natalie Diether, Faculty of Agriculture, Dalhousie University, Truro NS, Canada
Somayeh Kafaie, Department of Mathematics and Computing Science, SaintMary's University, Halifax NS, Canada
Lourdes Peña-Castillo, Departments of Computer Science and Biology, MemorialUniversity of Newfoundland, St. John's NL, Canada
Brent Robicheau, Department of Biology, University of Prince Edward Island,Charlottetown PE, Canada
Zoë Migicovsky, Department of Biology, Acadia University, Canada

**Presentation Overview:**
For over two decades, Bioinformatics User Groups (BUGs) have organized Bioinformatics, Computational Biology, and Data Science (BCBDS) community events across Canada to encourage local interaction, collaboration, and exchange of ideas. These established BUGs have successfully operated out of metropolitan areas. We sought to build on the success of existing BUGs and address the unique distribution of institutions across Atlantic Canada while simultaneously increasing accessibility to BCBDS resources within the region. We established BAtl – the Bioinformatics Atlantic Network in 2025. BAtl is the first dedicated BUG in Atlantic Canada, spanning ten institutions and servingstudents, researchers, educators and professionals. We designed BAtl programming to scale and reinforce bioinformatics community activities and expand to new local communities. We have implemented a turn-based model wherein the host institution rotates throughout the region to engage as many local participants as possible. Our hybrid meetings increase accessibility and continually grow our regional audience. Over our first five events we have engaged over 200 participants. Our model exemplifies that a distributed, turn-based seminar model is a viable strategy for engaging multiple institutions across a wider geographic region. Through this, we are building a thriving BCBDS community in Atlantic Canada by connecting local academic networks.

## A-110: A Generative Multi-Platform Framework for Isoform-ResolvedTranscriptomics at Single-Cell Resolution

**Authors List:**
**Jichen Wang**, McGill University, Canada
Michael Apostolides, McGill University, Canada
Ali Saberi, McGill University, Canada
Hamed Najafabadi, McGill University, Canada

**Presentation Overview:**

Alternative splicing is a key mechanism of gene regulation which expands transcript and protein diversity across cell types. Understanding isoform variation at single-cell resolution is essential for dissecting cellular heterogeneity in development and disease. However, accurate isoform quantification with single-cell RNA sequencing remains a major challenge. Short-read (SR) single-cell technologies suffer from strong 3?/5? positional biases and limited ability to distinguish isoforms, while long-read (LR) single-cell sequencing is constrained by low depth, high noise, and technical biases. To date, few computational frameworks effectively integrate these complementary data types for isoform-resolved single-cell analysis. We present Multi-Platform Aggregation and Quantification of Transcripts (MPAQT), a Bayesian generative model designed to integrate SR and LR single-cell RNA-seq data for accurate isoform quantification. MPAQT explicitly models platform-specific biases and incorporates machine learning–derived priors based on transcript sequence features to iteratively refine isoform quantification. Across simulations and experimental single-cell datasets, we show that MPAQT consistently outperforms state-of-the-art approaches at both gene and transcript levels. Applied to human embryonic stem cell differentiation into cortical neurons, MPAQT reveals that untranslated regions (UTRs) are highly correlated with cell-type-specific isoform usage. This framework enables a first-in-class isoform-resolved analysis of cellular heterogeneity across biological systems.

## A-112: Graph Neural Network Based Characterization of Human GutMicrobiomes

**Authors List:**
**Shida Sharif-Bakhtiar**, School of Computer Science, McGill University, Canada
Kai-Hsiang Chang, School of Computer Science, McGill University, Canada
Parham Ghasemloo Gheidari, School of Computer Science, McGill University, Canada
Jérôme Waldispuhl, School of Computer Science, McGill University, Canada
Roman Sarrazin-Gendron, Department of Computer Science, Université du Québec à Montréal, Canada

**Presentation Overview:**
Microbial communities that inhabit the human body play a significant role in the host's health. While we know that dysbiosis (lack of diversity) is linked to illness, its underlying mechanisms are poorly understood, and accurately predicting host illness from microbiome abundance profiles remains an open problem. Yet, the recent influx of tens of millions of genomic sequences brought by new sequencing techniques holds untapped potential. Several studies have been conducted to utilize standard evolutionary data such as phylogenetic trees; however, those cannot include reticulate evolution and horizontal gene transfers, common in microbes. We built microbial community networks, where each vertex corresponds to a microbe and edges are evolutionary relationships. Each of these networks is associated with host metadata. We apply graph neural networks (GNNs) to predict host status from these microbiome profiles. This graph classification approach allows more flexible representations of evolutionary distances than the state-of-the-art Convolutional Neural Networks (CNNs). GNNs can outperform CNNs on horizontal transfer-rich datasets, which are especially poorly represented by local neighborhoods. More importantly, being able to compare different graph representations provides opportunities to benchmark the contributions of various features on predictive power, which could unlock a path towards new insights regarding underlying mechanisms behind dysbiosis.

## A-114: Predicting Gene Expression Signatures from HistopathologyImages for Immunotherapy Biomarker Discovery

**Authors List:**
**Kewei Ni**, University of Toronto; University Health Network - Princess Margaret Cancer Centre, Canada
Kevin Wang, University Health Network - Princess Margaret Cancer Centre, Canada
Benjamin Haibe-Kains, University of Toronto; University Health Network - Princess Margaret Cancer Centre, Canada

**Presentation Overview:**
Background: Patient response to immune checkpoint blockade (ICB) remains heterogeneous, driven by complex tumor microenvironment (TME) interactions. Gene expression signatures are potent biomarkers for immunotherapy but limited by RNA sequencing cost and accessibility. Recent studies show that routine histopathology whole-slide images (WSIs) can predict molecular features, presenting opportunities to derive immunotherapy-relevant gene signatures from histopathology. Methods: We curated 52 published signatures spanning TME compositions and immune response groups. Using a pathology foundation model for feature extraction, we developed a multi-head attention-based multiple instance learning to predict signature scores from WSIs. Models were trained across 13 TCGA cohorts (n=5450) and validated in independent cohorts (n=1368). Results: Our Predictability Atlas characterizes which signatures are predictable from histopathology. We identified signatures with strong predictability (Pearson R > 0.5, p0.05), varying by cancer types. Eight signatures were consistently predictable across multiple cancers (R > 0.5, p0.05) as potential pan-cancer biomarkers. Attention heatmaps with cell segmentation confirm predicted signatures localize to biologically relevant regions. Ongoing work validates these histology-derived signatures for ICB response prediction in pan-cancer cohorts. Conclusion: This work links routine histopathology to molecular profiling, providing a framework for accessible, cost-effective immunotherapy biomarkers.

## A-116: Understanding the molecular machinery in agrochemicaldetoxification

**Authors List:**
**Aidan Jamieson**, York University, Canada
Mateus Peppinneli, York University, Canada
Laura Newburn, York University, Canada
Amro Zayed, York University, Canada

**Presentation Overview:**
The western honey bee, Apis mellifera, is the only honey bee in North America and is the primary managed pollinator in the world. The overuse of agrochemicals presents a current threat to honey bee populations. It can be difficult to disentangle and identify specific stressors that affect a particular colony. Studies directly linking exposure to agrochemicals and the elicited changes in gene expression are needed to improve our comprehension of the honey bee detoxification system. We are presenting the largest RNA-Seq study on honey bee toxicology, including the effect of 18 different agrochemicals from 5 classes, across 3 tissue types totalling 99 experimental contrasts. Using differential gene expression analysis and gene co-expression networks we show that each chemical produces both common and unique patterns of gene expression. Furthermore, detox gene regulation occurred in all tissue types and was present in all chemical classes, e.g., insecticides and herbicides. Alarmingly, this trend was prominent for amitraz and oxytetracycline, both used as treatments in commercial beekeeping. Our research addresses several questions in the field and provides important knowledge that can be mobilized to improve honey bee health and potentially the health of other native pollinators.

## A-118: Alignment-free copy number variation detection usingfull-text indices

**Authors List:**
**Clarence Todd**, University of Saskatchewan, Canada
Ian McQuillan, University of Saskatchewan, Canada
Lingling Jin, University of Saskatchewan, Canada

**Presentation Overview:**
Copy Number Variants (CNVs), such as insertions, deletions, and duplications, are DNA rearrangements that alter the copy number of DNA sequences across samples. Traditionally, CNV discovery relies on sequence alignment; however, scaling alignment to multiple whole genomes is computationally expensive without sacrificing sensitivity or avoiding the all-versus-all comparisons that compare every sequence to every other sequence, which increases the risk of missing valuable CNVs. Therefore, this work explores an alignment-free alternative for CNV discovery from full-text indices. Using a concatenated multi-genome full-text index consisting of the Burrows-Wheeler Transform and the suffix, longest common prefix (LCP), and document arrays, we propose a new method that discovers CNV variants by enumerating LCP intervals containing maximal repeats with varying copy counts across multiple genomes via an implicit and alignment-free all-versus-all comparison of the concatenated genomes. Using four simulated genomes created by introducing CNVs into a Hepatitis D reference genome, the approach successfully located every simulated tandem and interspersed duplication with near-optimal breakpoint resolution. Ongoing work is refining the method's effectiveness for detecting deletions and insertions, with additional experiments planned on real-world genome populations that provide accurate ground-truth data to fully validate the method's precision and recall metrics.

## A-120: Mapping the Genomic Niches of Transposable Elements: A Baseline-Controlled Resource Selection Framework Using Processed Pseudogenes

**Authors List:**
**Noah Zeidenberg**, University of Guelph, Canada
Karl Cottenie, University of Guelph, Canada
Brent Saylor, University of Guelph, Canada
Alexander Palazzo, University of Toronto, Canada
Rose Trappes, University of Bergen, Norway
Maegwin Bonar, University of Guelph, Canada
Stefan Linquist, University of Guelph, Canada

**Presentation Overview:**
Transposable elements (TEs) are ubiquitous mobile DNA sequences that account for approximately half of the human genome. Although TEs are major contributors to genomic plasticity, excess transposition can disrupt genomic stability and is associated with ageing and disease. Previous stochastic models indicate that TE insertions are spatially non-random but do not explicitly model known genomic architectural constraints, which influence TE insertion and persistence. Intronless pseudogenes (ILPs), nonfunctional gene copies generated by the retrotransposition machinery of the transposable element LINE-1, provide a mechanistic control for TE insertion and persistence biases. We introduce a comparative statistical framework that uses ILP distributions to define baseline expectations for TE insertional biases and applies resource selection functions to estimate feature-specific effect sizes for TE persistence relative to this baseline. Across genomic predictors including chromatin state, GC content, and histone marker density, the models reveal differences between TE and ILP persistence patterns at multiple spatial scales. By jointly modelling TEs, ILPs, and background genomic locations, we distinguish between shared mechanistic biases and TE-specific associations. This helps to provide an explanatory framework for how biochemical constraints and evolutionary forces interact to shape the genomic landscapes of transposable elements.

## A-122: Alternative Polyadenylation as a New Dimension of Biomarkers in Autism Spectrum Disorder

**Authors List:**
**Esther Jang**, The Hospital for Sick Children, Canada
Brett Trost, The Hospital for Sick Children, Canada
Maahil Arshad, The Hospital for Sick Children, Canada
Nickie Sarian, The Hospital for Sick Children, Canada

**Presentation Overview:**
Autism spectrum disorder (ASD) affects approximately 2% of children in Canada. Although heritability estimates are high (64–91%), indicating a strong genetic contribution, the molecular mechanisms remain incompletely understood due to substantial genetic and phenotypic heterogeneity. In the absence of pharmacological treatments for core symptoms, early detection combined with behavioral intervention remains the primary clinical strategy. We hypothesize that incorporating post-transcriptional regulatory variation beyond conventional genomic and transcriptomic analyses will improve ASD risk prediction. Alternative polyadenylation (APA) is a post-transcriptional mechanism that modulates 3?UTR length and influences mRNA stability, localization, and translation. While APA has emerged as a biomarker dimension in cancer, its role in ASD risk prediction remains largely unexplored. To address this, we identify ASD-associated APA features to enhance predictive modeling. We analyzed bulk RNAseq data from lymphoblastoid cell lines (LCLs) of 1,684 matched ASD case–control pairs from the Simons Simplex Collection (SSC), performing genome-wide quantification of 3?UTR usage and testing associations with phenotypic traits. Using mean group ASD case–control comparisons, we observed widespread 3?UTR shortening and lengthening genes enriched in ASD-related pathways, and aggregate shortening burden across ASD-associated genes showed opposite associations with ASD phenotypes, particular for intelligence measures between early-age and school-age ASD cases. These findings highlight APA as a developmentally sensitive regulatory layer contributing to ASD-related molecular variation and risk prediction.

## A-124: EpigeneticAgePipeline: an R package for comprehensiveassessment of epigenetic age metrics from methylationmicroarrays

**Authors List:**
Stanislav Rayevskiy, Department of Pathology and Laboratory Medicine, Western University, London, Ontario, Canada, Canada
Quinn Le, Department of Pathology and Laboratory Medicine, Western University, London, Ontario, Canada, Canada
Julia Nguyen, Department of Pathology and Laboratory Medicine, Western University, London, Ontario, Canada, Canada
Steven Q. Chen, Department of Pathology and Laboratory Medicine, Western University, London, Ontario, Canada, Canada
**Christina A. Castellani**, Department of Pathology and Laboratory Medicine, Western University, London, Ontario, Canada, Canada

**Presentation Overview:**
Epigenetic age is a biological age estimate based on nuclear DNA methylation patterns. Epigenetic clocks measure biological age by analyzing predictable changes in DNA methylation sites associated with aging. Here we introduce EpigeneticAgePipeline, an R package that streamlines the estimation of epigenetic age metrics including Horvath, Horvath skin and blood, Hannum, PhenoAge/Levine, GrimAge (V1 and V2), and DunedinPACE plus additional acceleration metrics based on all other clocks. In addition to the traditional clocks mentioned, the package also integrates a set of recently emerged epigenetic age clocks (PedBE, Wu, TL, BLUP, and EN). Quality control includes detection p-value filtering (sample- and probe-level), bead-count thresholds, and Illumina quality control intensity checks. EpigeneticAgePipeline supports Illumina Infinium methylation microarrays (HumanMethylation27, HumanMethylation450, HumanMethylationEPIC/EPICv2, and Human Methylation Screening Array). It offers functionalities including data preprocessing, normalization, cell count imputation, residual generation accounting for principal components and batch effects, and extensive visualizations for improved interpretability. EpigeneticAgePipeline provides an integrated workflow from raw data to advanced statistical analyses and visualizations, improving usability over existing tools. Future updates will include emerging epigenetic age measures to maintain relevance in this evolving field.

## A-126: Exploring Neurodevelopmental Impacts of SETD2 Mutations Through Bulk and Single-Cell Multi-omics

**Authors List:**
**Parsa Seyfourian**, Graduate Program in Bioinformatics, UBC, Vancouver Canada; BC Cancer Research Centre, Vancouver Canada, Canada
Emily Yeh, BC Cancer Research Centre, Vancouver Canada; Department of Biochemistry and Molecular Biology, UBC, Vancouver Canada, Canada
Lauren Blume, BC Cancer Research Centre, Vancouver Canada; Graduate Program in Neuroscience, UBC, Vancouver Canada, Canada
Pariya Azarafshar, BC Cancer Research Centre, Vancouver Canada, Canada
Yongjin Park, BC Cancer Research Centre, Vancouver Canada; Department of Pathology and Laboratory Medicine, UBC, Vancouver Canada, Canada
Carol Chen, BC Cancer Research Centre, Vancouver Canada; Department of Biochemistry and Molecular Biology, UBC, Vancouver Canada, Canada

**Presentation Overview:**
Neurodevelopmental disorders (NDDs) affect 1 in 10 children in Canada and can encompass cognitive, behavioural, and motor impairments. Despite their prevalence, the biological mechanisms underlying these disorders remain poorly understood. Mutations in SETD2, a chromatin modifier, are linked to severe NDDs and paediatric high-grade gliomas. Different SETD2 mutations can result in contrasting conditions, such as Luscan-Lumish Syndrome (brain overgrowth and mild disability) and Rabin-Pappas Syndrome (brain underdevelopment and profound disability). How SETD2 mutations alter brain development at the cellular level remains unclear. We used cerebral organoids, stem-cell derived models of human brain development, to study the impact of SETD2 mutations on neurodevelopment. Bulk, scRNA, and epigenomics sequencing were utilized to analyze gene expression and chromatin state at global and single-cell resolution. By comparing SETD2 mutant organoids with unedited controls, we uncover how these mutations affect cellular composition, differentiation, and regulatory pathways. Preliminary observations show striking phenotypic differences in organoids including drastic morphological differences, changes in neuronal-progenitor cell identities, and emergence of non-neuronal cell types. Using integrated multi-omics approaches, this project aims to identify disrupted regulatory networks and signaling pathways, informing future strategies for genetic screening and therapies.

## A-128: Rethinking Transposable Element Expression Analysis

**Authors List:**
**Elspeth Smith**, University of Guelph, Canada
Yan Yan, University of Guelph , Canada
Jasmin Lalonde, University of guelph, Canada

**Presentation Overview:**
Transposable elements (TE) are evolutionarily ancient, mobile DNA sequences capable of self-replication. In humans, TE activity is largely restricted to the brain where they play a role in healthy brain development. However, TE dysregulation has also been implicated in the development of several complex mental health disorders. RNA-seq analysis is commonly employed to study TE activity that generally includes mapping RNA reads to a reference genome. However, high rates of polymorphisms and the presence of non-reference TEs (nrTE) hinder accurate analysis of TE expression, though the extent and impact of this has not been fully explored. In this work, the telomere-to-telomere (T2T) genome assembly has been employed to simulate RNA-seq reads from nrTEs. Simulated reads are mapped to the hg38 human reference genome, mirroring a typical RNA-seq analysis pipeline. Output counts are compared to the simulated ground truth. When analysis was performed at the family level the presence of nfTEs resulted in minimal loss of accuracy, whereas locus-level quantification resulted in a significant reduction of accuracy. This suggests that current pipelines for loci specific TE expression analysis are not as accurate as previously described, pointing to the need to develop pipelines that effectively address the presence of nrTEs.

## A-130: SOULCAP Template-Guided Label Transfer for StandardizedCytometry Population Annotation

**Authors List:**
**Leon Li**, McGill University, Canada
Justin Meskas, Dotmatics, Canada
Ryan Brinkman, Dotmatics, Canada
Jérôme Waldispühl, McGill University, Canada

**Presentation Overview:**
A major challenge to interpretable and scalable flow cytometry data analysis is the lack of globally standardized cell-population annotations. The current reporting standard, MIFlowCyt, is insufficient and often poorly enforced to ensure ontology-aligned population identification. Consequently, decades of qualitative, lab-specific naming conventions have produced inconsistent labels that limit cross-study data integration and hinder modern machine-learning development. SOULCAP, a non-profit community initiative, addresses this problem by generating expert-curated reference bivariates that quantitatively define immunological populations. Here, we present a quantitative label-assignment algorithm, Vis-Align, that automatically annotates both historical and newly generated datasets by aligning user data to SOULCAP reference bivariates. These references define populations through essential immunophenotypic marker combinations and gating logic. Rather than relying on overly complex and computationally heavy models, Vis-Align encodes simple visual and structural cues that humans use during alignment. We evaluated multiple assignment strategies across bivariate and multidimensional approaches. Vis-Align demonstrated strong generalizability, producing consistent and high-confidence population assignments even when user data differed substantially from the reference, achieving a strong balance of speed, robustness, and accuracy. By enabling standardized and ontology-aligned population annotations, this work offers a foundational resource for reproducible cytometry data analysis, large-scale data harmonization, and next-generation AI development in immunology.

## A-132: Investigating the mediating role of metabolites in mtDNA-CN and nuclear epigenome driven cardiovascular disease risk using a large-scale longitudinal cohort

**Authors List:**
**Rachel Barboza**, 1. Pathology and Laboratory Medicine, Western University, Canada
Phyo Win, 1. Pathology and Laboratory Medicine, Western University, Canada
Steven Chen, 1. Pathology and Laboratory Medicine, Western University, Canada
Osvaldo Espin-Garcia, 1. Epidemiology and Biostatistics, Western University 2. Biostatistics and Schroeder Arthritis Institute, University Health Network 3. Dalla Lana School of Public Health and Department of Statistical Sciences, University of Toronto, Canada
Christina A. Castellani, 1. Pathology and Laboratory Medicine, Western University 2. Children's Health Research Institute, London Health Sciences Centre Research Institute 3. Genetic Medicine, Johns Hopkins University, Canada

**Presentation Overview:**
Variation in mitochondrial DNA Copy Number (mtDNA-CN), a proxy for mitochondrial function, is associated with complex diseases. mtDNA communicates with the nucleus via epigenetic mechanisms that regulate nuclear gene expression, however, the role the metabolome plays in mediating these relationships remains unknown. mtDNA-CN can be derived from nuclear-mitochondrial ratios, but estimates are seldom available and fail to integrate multi-omic signatures, demonstrating the need for novel methods that improve mtDNA-CN estimation. The Canadian Longitudinal Study of Aging is a human cohort with matched genomic, metabolomic, and epigenomic data. Baseline mtDNA-CN estimates were tested for association with 1,022 metabolites to assess if metabolites mediate the relationship between mtDNA-CN and nuclear DNA (nDNA) methylation in disease. Further, a heterogenous transfer learning (HTL) model was applied to multi-omic data to impute and improve mtDNA-CN prediction. Ninety-three metabolites were associated with mtDNA-CN, including the epigenome-modifying metabolites S-adenosylhomocysteine and alpha-ketoglutarate (N=9,375, FDR=4.25×10-3, 2.93×10-3). mtDNA-CN associated metabolites, including cholesterol and sphingomyelins, were associated with myocardial infarction risk (n=57, FDR0.05). The HTL model using disparate omic datasets demonstrated high predictive power. These findings suggest mtDNA variation remodels the nuclear epigenome via methionine cycle metabolites, a regulator of nDNA methylation, which contributes to increased cardiovascular risk.

## A-134: Host Finder: computational detection of host-microbe interactions by large-scale data-mining of the Sequence Read Archive

**Authors List:**
**Shyan Mascarenhas**, University of Waterloo, Canada
Harold Hodgins, University of Waterloo, Canada
Andrew Doxey, University of Waterloo, Canada

**Presentation Overview:**
The Sequence Read Archive is the largest public sequencing repository and represents an unparalleled record of global biodiversity. Despite its scale, it remains underutilized for identification of host–microbe associations. We present HostFinder, a computational framework to infer host–microbe associations across ~30 million SRA accessions. Taxonomic profiles generated using STAT were aggregated to quantify host and microbial abundances per dataset, and association strength was measured using k-mer abundance thresholds and a log-odds co-occurence score. Using curated eukaryote–bacteria and eukaryote–virus interactions from PHI-Base and Virus-Host DB, we tested whether co-occurrence scores distinguish true interactions from false ones. Using the co-occurrence score alone, 81% of virus–host (n = 43), 56% of bacterial pathogen–host (n = 157), and 61% of bacterial commensal–host (n = 100) interactions were correctly identified, with 12% FDR. Scaling across the full SRA, 6 billion host–microbe pairs were screened, identifying 3.56 million high-confidence interactions (log-odds > 3, FDR 1%). As an example of host–virus interactions detected by our pipeline, we identified novel Partitiviridae–insect associations, independently corroborated through viral genome assembly and phylogenetic analysis. HostFinder shows that global co-occurrence analysis can recover known and novel associations at scale.

## A-136: Assessing the impact of superinfection and recombination onphylodynamic inference by multi-level simulation

**Authors List:**
**Delicia Wong**, Western University, Canada
William Wang, Western University, Canada
Art Poon, Western University, Canada

**Presentation Overview:**
Phylodynamic methods estimate key epidemiological parameters, including the basic reproduction number (R?), from viral sequence data by linking phylogenetic tree structure to transmission dynamics. These approaches typically assume that sequences are related by a single phylogenetic tree. However, superinfection allows divergent viral lineages to co-infect hosts and undergo recombination, generating discordant evolutionary histories across the genome that cannot be represented by a single tree and may bias inference. We quantified the impact of superinfection and recombination on R? estimates using a multi-level simulation framework. Transmission trees and nested within-host phylogenies were simulated under a susceptible–infected–removed model with known R? values (1.96 for HIV-1; 3.38 for SARS-CoV-2), and phylodynamic inference was performed using the birth–death SIR model in BEAST2. Recombination was introduced via ancestral recombination graphs, and sequence alignments (100 tips) were simulated and analyzed. Baseline simulations accurately recovered R? (RMSE=0.06 for HIV-1; 0.12 for SARS-CoV-2). With increasing recombination, reconstructed trees became increasingly star-like, and R? estimates were significantly biased upward by +0.012 (HIV-1) and +0.041 (SARS-CoV-2) per 10 breakpoints. These findings demonstrate that recombination can substantially inflate phylodynamic estimates of transmission potential, highlighting key limitations of standard tree-based methods for recombining viruses.

## A-138: Graph-Based Genomic Context Analysis Reveals Bacterial Defence System Architecture

**Authors List:**
**Leila Moghaddam Moghanloo**, Department of Computer Science, University of Saskatchewan,Saskatoon, Saskatchewan, Canada, Canada
Chantel N. Trost, Genetics and Genome Biology Program, The Hospital for SickChildren (SickKids), Toronto, Ontario, Canada, Canada
Lingling Jin, Department of Computer Science, University of Saskatchewan,Saskatoon, Saskatchewan, Canada, Canada

**Presentation Overview:**
Bacterial defence systems against phages and mobile genetic elements are highly diverse and evolve rapidly, limiting similarity-based approaches for identifying defence genes and inferring functional relationships. Functionally related genes often cluster within genomic neighbourhoods, where conserved organizational patterns may complement sequence-based methods, particularly in highly variable regions. Despite this potential, contextual information remains underutilized in computational inference of defence gene function and evolution. We develop a network-based approach to analyse genomic neighbourhood context of defence-associated genes across bacterial genomes. A locus is defined as a focal gene with its ten upstream and downstream genes, approximating operon and defence island organization. Genes are annotated for contextual features, including transposases, integrases, and defence-related markers. These annotations are encoded in a locus–feature matrix used to construct a bipartite graph, linking loci to features, which is transformed into a locus-similarity network, where edges reflect shared neighbourhood composition. Community detection identifies groups of loci with similar contextual organization. Preliminary analysis reveals non-random clustering of defence-associated loci enriched for mobile-genetic-element–related features, consistent with known patterns. Expansion across diverse bacterial phyla is expected to uncover novel defence-associated modules for experimental validation. This establishes genomic neighbourhood modelling as a strategy for functional inference in evolving bacterial defence systems.

## A-140: ChargeHydro-Net: A Physicochemically-Informed Deep LearningArchitecture for Antimicrobial Peptide Prediction

**Authors List:**
**Rishik Kondadadi**, University of Minnesota, Eastview High School, United States
Dan McCreary, Kelly-McCreary Associates, United States

**Presentation Overview:**
Antimicrobial resistance threatens to cause 10 million deaths annually by 2050, making the discovery of novel antimicrobial peptides (AMPs) an urgent priority. While deep learning models for AMP prediction typically rely on sequence information alone, AMPs are fundamentally defined by their physicochemical properties: cationic charge drives electrostatic attraction to anionic bacterial membranes, while amphipathic hydrophobicity enables membrane insertion. We present ChargeHydro-Net, a dual-stream neural network that explicitly incorporates residue-level charge, hydrophobicity, and molecular mass alongside learned sequence embeddings. The sequence stream processes amino acids through a bidirectional LSTM with self-attention, while a parallel property stream encodes fixed physicochemical features. A cross-stream attention mechanism allows the model to query biophysical context at positions identified as critical by sequence patterns, integrated via a gated fusion layer. Evaluated on the AMPlify benchmark (9,014 sequences), ChargeHydro-Net achieves an AUROC of 0.9696 and an MCC of 0.8491, outperforming ESM-2 fine-tuning and standard BiLSTM baselines. These results demonstrate that encoding known biophysical priors directly into the architecture captures AMP-relevant biology that sequence-only models overlook.

## A-142: Conditional Genetic Interaction Landscapes of Yeast WGDParalogs Across Diverse Metabolic Environments

**Authors List:**
**Mohammadreza Yasemi**, Concordia University, Canada
Brittany M Greco, Concordia University, Canada
Elena Kuzmin, Concordia University, McGill University, Canada

**Presentation Overview:**
Whole-genome duplication (WGD) in Saccharomyces cerevisiae (~100 million years ago) generated 551 duplicate gene pairs, and evidence suggests angiosperm-driven, sugar-rich fermentation niches contributed to their retention. Here, we profile 79 WGD paralog pairs (158 single and 79 double gene deletion mutants) characterized by sparse digenic and trigenic interactions in rich media and quantify fitness across 30 combinations of different carbon and nitrogen environments using automated time-lapse colony imaging. After correcting spatial, plate and batch effects, we compute area-under-the-curve growth metrics and use statistical models to identify gene–environment (G×E) and gene–gene–environment (G×G×E) effects that distinguish functional divergent and redundant paralogs. Preliminary analysis shows a substantial fraction of paralog mutants display condition-specific fitness defects. To score G×G×E, we adapted the ?-SGA trigenic interaction quantitative framework by treating the environment as the third axis. Conditional single- and double-mutant fitness defects are most consistently observed under maltose, galactose and raffinose-containing conditions, which are important for ripening and stress response. In contrast, glucose, fructose and sucrose conditions show fewer severe defects. Overall, the defect burden peaks in environments that require substantial metabolic rewiring for alternative carbon utilization, consistent with these niches being where paralog specialization and buffering are most strongly unmasked.

## A-144: Sustainability, Intellectual Property, and theComprehensive Antibiotic Resistance Database

**Authors List:**
**Andrew McArthur**, McMaster University, Canada
Brian Alcock, McMaster University, Canada
Amogelang Raphenya, McMaster University, Canada

**Presentation Overview:**
Increasing rates of antimicrobial resistance (AMR), i.e., drug-resistant infections, is a direct threat to modern medicine. In Canada, approximately 26% of infections in 2018 were resistant to the drugs generally used to treat them. Globally, current rates of AMR and predictions for increased rates of DRI are equally dire. Yet, antimicrobial resistance in human populations does not occur in isolation but instead occurs within the broader context of antimicrobial use in agriculture and the pollution of natural environments. Canada has instituted a Federal Action Plan on Antimicrobial Resistance and Use, of which one focus is surveillance of AMR. The Comprehensive Antibiotic Resistance Database (card.mcmaster.ca) is a made-in-Canada database and software platform used around the world for genomic surveillance of antimicrobial resistance genes (ARGs) by applying diverse data science approaches from genomics to natural language processing to protein language models. Although an academic enterprise, CARD by design does not rely on grant funding for sustainability and support of professional biocurators and bioinformaticians. CARD's data and resources are free for academics and public health agencies but require a license for private sector use, requiring a careful balance of academic research, service to the broader community, dedication to open source, and commercialization.

## A-146: The Canadian Genomic Data Commons (CGDC): A NationalInfrastructure for Genomic Data Sharing

**Authors List:**
**Erika Frangione**, Mount Sinai Hospital, Sinai Health, Canada
Jordan Lerner-Ellis, Mount Sinai Hospital, Sinai Health, Canada
Navneet Aujla, Mount Sinai Hospital, Sinai Health, Canada
Heidi Rehm, Broad Institute, United States
Marc Fiume, DNAstack, Canada
Vincent Ferretti, Université de Montréal, Canada
Yann Joly, McGill University, Canada
Steven Jones, BC Cancer, Canada
Patrick Frosk, University of Manitoba, Canada
Sherryl Taylor, University of Alberta, Canada
Kym Boycott, University of Ottawa, Canada

**Presentation Overview:**
Clinical genome sequencing (GS) generates large data volumes with prohibitive storage costs, and remains largely inaccessible to researchers and clinicians. The Canadian Genomic Data Commons (CGDC) will use high-performance computing (HPC) to securely transfer genomic data from healthcare settings into a data sharing ecosystem. This is especially urgent for rare diseases (RDs), which afflict >3 million Canadians. The CGDC will develop three core facilities to create sustainable, federated, and secure digital infrastructure for data sharing across Canada: 1. The Canadian Open Genetics Repository (COGR); 2. Canadian genome aggregation database (gnomAD-Canada); and, 3. Tools for RD researchers. 1) COGR will centralize the standardization and sharing of variant data from diagnostic laboratories through laboratory-specific automated data transfer pipelines, variant discrepancy reporting, and submission of consensus interpretations to COGR and ClinVar. 2) A Canadian instance of the gnomAD browser will be established using aggregated allele frequencies from Canadian large-scale GS projects. 3) A Canadian deployment of the seqr platform on HPC4Health will support gene–disease discovery, alongside a dynamic consent portal developed in collaboration with the Pan-Canadian Genome Library (PCGL). The CGDC will expand HPC research capacity, improve variant interpretation and quality assurance, and maximize the value of Canadian clinical and research cohorts.

## A-148: GenPipes: a workflow manager for accessible andreproducible bioinformatics

**Authors List:**
**Mareike Janiak**, McGill University / Canadian Centre for Computational Genomics, Canada
Paul Stretenowich, McGill University / Canadian Centre for Computational Genomics, Canada
Robert Eveleigh, McGill University / Canadian Centre for Computational Genomics, Canada
Jean-Michel Garant, McGill University / Canadian Centre for Computational Genomics, Canada
Alexa Li Kam Wa, McGill University / Canadian Centre for Computational Genomics, Canada
Jose Hector Galvez Lopez, McGill University / Canadian Centre for Computational Genomics, Canada
Guillaume Bourque, McGill University / Canadian Centre for Computational Genomics, Canada

**Presentation Overview:**
GenPipes is an open-source Python-based workflow management system offering 10 pre-built bioinformatics pipelines. Its features include smart restarts, easy configuration, support for multiple job schedulers and cloud computing, detailed logs and documentation. An interactive command-line wizard provides guidance for beginners, while ticket support is available to all users. GenPipes is integrated with the clusters of the Digital Research Alliance, where it can be loaded as a module and accesses a central software stack, so users do not need to handle any installations. Pre-built pipelines are available for RNA-seq, long-read sequencing, and WGS, among others, and include protocols such as paired tumour/normal analysis. This protocol has been used to analyze more than 4400 cancer cases for the Marathon of Hope project in Québec, illustrating its scalability and applicability to large projects. Continuous improvements are made to pipelines in response to user feedback and internal benchmarking. Current development goals include better support for the T2T human reference, enhancing of actionable reporting for somatic variants, and improving pipelines for long-reads. GenPipes is of interest to a broad user base, as it both lowers the barrier of entry to bioinformatic analysis and ensures research reproducibility, crucial for large, multi-year projects.

## A-150: Chicken or Egg: did immune ligands evolve before their receptors?

**Authors List:**
**Colin O'Dwyer**, Ottawa Hospital Research Institute, Canada
David Cook, Ottawa Hospital Research Institute, Canada
Michele Ardolino, Ottawa Hospital Research Institute, Canada

**Presentation Overview:**
Due to its short cytoplasmic tail, PD-L1 has historically been considered merely an inert ligand for its canonical receptor PD-1. However, emerging evidence illustrates that PD-L1 possesses a range of cell-intrinsic functions independent of PD-1. Given these intrinsic roles in the absence of its receptor, we investigated whether these functions represent ancestral roles from before PD-L1 evolved as a ligand for PD-1. We developed an evolutionary analysis pipeline that aligns human proteins to animal proteins across 24 curated databases spanning Metazoa. The pipeline identifies the best aligning animal proteins and derives information about their conserved domains, synteny, and structural similarities to the human protein of interest. Using this approach, we deduced the evolutionary conservation of PD-L1 and PD-1. Our analysis suggests that PD-L1 appears earlier in evolutionary time than PD-1, potentially explaining why PD-L1 retains cell-intrinsic effects even in the absence of its receptor. Additionally, we found evidence suggesting that many immune ligands evolved before their cognate receptors, highlighting an intriguing broader evolutionary pattern. Positioned at the intersection of comparative immunology and evolutionary biology, this work addresses gaps in our fundamental understanding of immune proteins and provides evolutionary context for the multifunctional nature of immune checkpoint molecules.

## A-152: Graph Attention Networks for Chromosome-Level Modeling of cfDNA Methylation in Cancer Detection

**Authors List:**
**Lin Zhao**, Algoma University, Canada
Vanessa Giuliano, Algoma University, Canada
Yan Yan, University of Guelph, Canada
Wenjun Lin, Algoma University, Canada
Ping Luo, Algoma University, Canada

**Presentation Overview:**
Recent advances in cell-free DNA (cfDNA) methylation profiling have enabled minimally invasive cancer detection using plasma samples, providing a promising alternative to conventional tissue-based diagnostics. Existing approaches typically combine cfDNA methylation signals with statistical feature selection and machine-learning classifiers to identify cancer samples. However, these methods often restrict learning to top-ranked differentially methylated regions (DMRs), potentially overlooking informative signals, and rely on fixed-size genomic bins that fail to capture DMRs with their true variable lengths. As a result, complex methylation patterns and long-range genomic dependencies at the chromosome scale are insufficiently modeled. To address these limitations, we propose a graph-structured modeling framework that explicitly incorporates genomic locations and relationships among chromosomal bins, enabling a biologically informed representation of cfDNA methylation data. Chromosomes are modeled as a graph, with nodes representing genomic bins and edges encoding positional and contextual relationships. A Graph Attention Network (GAT) is applied to automatically learn methylation patterns while simultaneously identifying informative subgraphs corresponding to DMRs. The proposed method is evaluated on 628 cell-free methylated DNA immunoprecipitation and high-throughput sequencing samples, achieving an AUC of 0.949 and outperforming previously reported methods. These results highlight the potential of graph neural networks for chromosome-level cfDNA analysis and cancer detection.

## A-154: Dual curation of diploid genomes improves assembly quality

**Authors List:**
**Erick Duarte**, The Rockefeller University, United States
Kirsty McCaffrey, The Rockefeller University, United States
Jack Medico, The Rockefeller University, United States
Giulio Formenti, The Rockefeller University, United States
Delphine Lariviere, The Rockefeller University, United States
Erich Jarvis, The Rockefeller University, United States

**Presentation Overview:**
Genome assembly algorithms face challenges such as high repetitiveness when piecing together sequence data from non-model genomes, often producing fragmented assemblies with errors. Correcting these misassemblies requires manual curation, including sequence breaks, joins, and reorientations at the contig and scaffold levels. which substantially improves assembly quality. Traditionally, curation has focused on a single haplotype or a pseudo-haploid representation of diploid genomes, potentially reducing haplotypic diversity and masking true variation. As fully phased assemblies become the standard, we present a “dual manual curation” methodology that involves simultaneous curation of both haplotypes of a diploid organism using a single Hi-C map. This approach minimizes structural variation errors, increases the accuracy of both haplotypes, improves the identification and curation of sex chromosomes, and yields two high-quality near-complete genomes. It also simplifies the resolution of misphasing events and streamlines the manual curation process. To support this approach, we have automated pre- and post-dual curation workflows in the Galaxy workspace. The automated workflows streamline the data processing, facilitate the data generation for curation decisions, and are less computationally demanding than the independent curation of both haplotypes. Dual curation has now become the main curation approach for the Vertebrate Genomes Project and Earth Biogenome Project.

## A-156: A systematic assessment of the effect of datarepresentation and machine learning methods on sampleclassification based on microbiota composition

**Authors List:**
**Paniz Taghipour**, Memorial University of Newfoundland, Canada
Lourdes Pena-Castillo, Memorial University of Newfoundland, Canada

**Presentation Overview:**
Microbiota composition is associated with various phenotypes and diseases, making it an increasingly attractive target for diagnostic tools and therapeutic interventions. However, to achieve this, microbiome studies require reproducible, scalable workflows addressing compositional data constraints while balancing model interpretability with classification performance. We developed a Nextflow pipeline for 16S rRNA amplicon data analysis following best practices for microbiome analysis and evaluation of machine learning (ML) models. Our pipeline generates four commonly-used data transformations: genus-level relative abundances, presence–absence encodings, centered log-ratio transformations, and phylogenetic isometric log-ratio balances incorporating evolutionary relationships. Each of these data representations is used as input to a Random Forest classifier and a Neural Additive Model to generate models for sample classification. Each data transformation / ML method combination is evaluated using five independent datasets spanning a wide range of sample sizes ([50,1000]), hosts (human, dog, murre) and associated phenotype. We compare feature transformation methods and machine learning classifiers using the area under the precision–recall curve (AUPRC). Preliminary results indicate clear sample size dependencies and transformation-specific performance patterns. In particular, Neural Additive Models require substantially larger datasets than Random Forest to achieve equivalent predictive performance. Our results will provide evidence-based guidance for methods selection in microbiome studies.

## A-158: The Impact of Strain Selection on Reverse Vaccinology basedAntigen Candidate Identification

**Authors List:**
**Justin Schonfeld**, Public Health Agency of Canada, Canada
Shaun Kernaghan, Public Health Agency of Canada, Canada
Joenel Alcantara, University of Calgary, Canada
Erin Brown, University of Calgary, Canada

**Presentation Overview:**
Salmonella enterica is a major foodborne pathogen impacting both humans and animals, imposing heavy economic and health burdens globally. Many non-typhoidal serovars of Salmonella sp. are zoonotic in nature existing largely in animal reservoirs. Subunit vaccines, vaccines composed of multiple antigenic pieces of a pathogen, offer the potential to provide safe long lasting immunity across multiple serovars. In recent years vaccine development has been accelerated through the use of reverse vaccinology, the application of computational methods to a single whole genome sequence in order to identify promising antigen candidates. However, increasing use of WGS for pathogen surveillance has both highlighted the genetic diversity of many zoonotic Salmonella serovars and created the opportunity to investigate the impact of that genetic diversity on potential antigenicity of individual genes. Using a custom reverse vaccinology pipeline we analyzed 322,379 Salmonella enterica genomes. From those genomes we identified two sets of candidate genes: those found in all Salmonella enterica serovars and those found in zoonotic serovars associated with chicken. We found that strain selection can have an impact on predicted antigenicity, compared existing machine learning approaches to predicting antigenicity, and developed a weighted ranked list of candidates for use in the development of future vaccines.

## A-160: Towards interoperable and FAIR water-based genomicsurveillance: the iMicroSeq contextual data specificationpackage

**Authors List:**
**Charlie Barclay**, Centre for Infectious Disease Genomics and One Health, Faculty of Health Sciences, Simon Fraser University, Canada
Rhiannon Cameron, Centre for Infectious Disease Genomics and One Health, Faculty of Health Sciences, Simon Fraser University, Canada
Nima Tehrani, Brinkman Lab, Department of Molecular Biology and Biochemistry, Simon Fraser University, Canada
Baofeng Jia, Brinkman Lab, Department of Molecular Biology and Biochemistry, Simon Fraser University, Canada
Erin E Gill, Brinkman Lab, Department of Molecular Biology and Biochemistry, Simon Fraser University, Canada
Justin Richardsson, Ontario Institute for Cancer Research, Canada
William Wl Hsiao, Centre for Infectious Disease Genomics and One Health, Faculty of Health Sciences, Simon Fraser University, Canada
Fiona Sl Brinkman, Brinkman Lab, Department of Molecular Biology and Biochemistry, Simon Fraser University, Canada
Mélanie Courtot, Ontario Institute for Cancer Research and Department of Medical Biophysics, Univers

**Presentation Overview:**
Water-based environmental DNA (eDNA) surveillance is an increasingly used tool to monitor pathogens, antimicrobial resistance, biodiversity, and ecosystem change across aquatic environments. However, fragmented data management practices limit interoperability, cross-sectoral integration, and long-term reuse. As part of the Genome Canada funded iMicroSeq project, we developed a water-based microbial genomics data specification to standardise contextual data to facilitate interoperable water-based genomics data across surveillance and research programs. The specification was developed using LinkML and aligned with OBO Foundry principles, MIxS, and ISO standards. It extends the international PHA4GE wastewater model to support diverse aquatic matrices, including freshwater, marine, drinking water, and built systems. The ontology-driven schema harmonizes metadata across sampling context, environmental parameters, eDNA collection, laboratory processing, sequencing, and bioinformatics workflows. Controlled vocabularies and ontology mappings enable semantic validation, machine-readable data exchange, and interoperability. Implemented through DataHarmonizer templates, the model enables structured submission, standardised curation, and reproducible data workflows. Beyond standardisation, the specification serves as the architectural framework for the new iMicroSeq Data Portal, directly informing its data model, validation rules, and interoperability design. Comprising over 250 harmonized fields and 900+ controlled terms, this infrastructure strengthens FAIR data stewardship and supports scalable, cross-domain water-based genomics surveillance in Canada and beyond.

## B-001: Tandem Repeat Expansions Drive Genomic Instability andClinical Outcomes Across Cancers

**Authors List:**
**Alexander Turco**, Princess Margaret Cancer Research Centre, Canada
Sushant Kumar, Princess Margaret Cancer Research Centre, Canada

**Presentation Overview:**
Repetitive DNA sequences constitute over half the human genome and play critical roles in shaping its structure, stability, and regulation. Yet, much of cancer genomics remains focused on the small fraction of the genome that encodes proteins, overlooking the vast repetitive landscape that may underlie genomic instability. While microsatellite instability has been well characterized in specific cancers, the broader landscape of tandem repeat (TR) expansions across tumour genomes remains poorly understood. By systematically profiling TR expansions across 16,000 patients from 38 different cancer types, we reveal distinct loci of repeat instability, some recurrent and shared across many cancer types, others uniquely enriched within specific tumours. Although most expansions arise within noncoding and intronic regions, they consistently overlap candidate cis-regulatory elements (cCREs) and genes linked to disease and transcriptional control, suggesting functional consequences within what has been notoriously defined “junk” DNA. Enrichment analyses highlight potential associations between repeat expansions, transcriptional regulation, and chromatin structure. Additionaly, variation in expansion burden correlates with differences in patient survival, revealing potential clinical relevance. Collectively, these findings reveal a previously overlooked regulatory layer of genomic instability and establish a foundation for understanding how tandem repeat Expansions contribute to oncogenesis.

## B-003: Genetic and Neural Pathways to Distressing Psychotic-LikeExperience Subtypes in Early Adolescence

**Authors List:**
**Benson Ku**, Emory, United States
Ying Xu, Emory, United States
Ella Arrant, Emory, United States
Cope Feurer, UNC, United States
Yinxian Chen, Emory, United States
Shalaila Haas, Mount Sinai, United States
Anke Huels, Emory, United States

**Presentation Overview:**
Objective: Delusions and hallucinations are traditionally grouped as positive psychotic symptoms, yet evidence suggests they may arise from distinct genetic and neural mechanisms. This study aimed to characterize trajectories of distressing paranoid and hallucinatory psychotic-like experiences (PLEs) in early adolescence and examine how polygenic risk scores for schizophrenia (PRS-SCZ) influence these trajectories through cortical thickness in medial prefrontal cortex (mPFC) subregions. Methods: This longitudinal study followed 9–10-year-old participants from the Adolescent Brain Cognitive Development Study. We derived PRS-SCZ from genome-wide association study summary statistics and computed cortical thickness of mPFC regions from structural MRI. We identified trajectories of distressing paranoid and hallucinatory PLEs across 4 years. Structural equation modeling tested associations among PRS-SCZ, mPFC thickness, and distressing PLEs. Results: Among 11,666 participants, 2,411 reported paranoid PLEs and 3,934 reported hallucinatory PLEs. Higher PRS-SCZ was associated with reduced right frontal pole thickness (?=?0.028, 95% CI=?0.045, ?0.011), which in turn was associated with greater odds of distressing paranoid PLEs (adjusted odds ratio [aOR]=0.95, 95% CI=0.91, 0.99). Higher PRS-SCZ was associated with greater odds of distressing hallucinatory PLEs (aOR=1.05, 95% CI=1.01, 1.09). Conclusions: Genetic liability for schizophrenia may shape trajectories of both distressing paranoid and hallucinatory PLEs, although neurobiological mechanisms may differ.

## B-005: From Chaos to Convergence: Modelling Gut Virome Diversity, Dynamics, and Function Across 12 Infant Cohorts

**Authors List:**
**Michael Shamash**, McGill University, Canada
Corinne F. Maurice, McGill University, Canada

**Presentation Overview:**
The early-life gut virome is highly dynamic, yet extreme inter-individual variability and batch effects often obscure universal developmental patterns. We conduct a computational meta-analysis of 12 infant cohorts (1,893 samples from 1,089 infants) to model gut virome maturation over the first three years of life. Statistical modelling was used to control for technical and biological confounders. Our results revealed distinct diversity patterns: viral richness increased for seven months before plateauing, while stable Shannon diversity indicated a loss of community evenness as the virome matures. We quantified ""virome developmental velocity"" (rate of change between sequential samples), which was high in early life but significantly decreased over time. This was mirrored by community-wide convergence (lower beta dispersion) as viromes became more similar with age. Functionally, this transition involved a significant decrease in temperate phages and shifts in phage-encoded auxiliary metabolic genes (AMGs), which can modulate host metabolism. Finally, a random forest regression model accurately predicted infant age during the early, dynamic phase, but lost its predictive power after 20 months. Overall, the infant gut virome transitions from a highly dynamic, individualized state to a stable, functionally convergent community.

## B-007: Benchmarking CUT&RUN analysis procedure using motifenrichment

**Authors List:**
**Luomeng Tan**, University of Toronto, Canada
Coby Viner, University of Toronto, Canada
Michael Hoffman, University of Toronto, Canada

**Presentation Overview:**
Cleavage Under Targets and Release Using Nuclease (CUT&RUN) maps the genome-wide locations of chromatin-associated proteins, including transcription factors. Identifying sequence-specific transcription factor binding sites plays a critical role in understanding gene regulation. CUT&RUN provides an improved alternative to chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) for this purpose. Existing computational tools primarily target ChIP-seq data and often perform suboptimally on CUT&RUN because of its distinct fragment length and cleavage characteristics. In this project, we focus on improving preprocessing strategies for CUT&RUN, such as fragment length filtering and spike-in calibration, that influence peak detection. We designed a benchmarking method to evaluate peak calling procedures for CUT&RUN data and the effects of varying preprocessing approaches. We benchmarked the two most widely-used peak callers, MACS2 and SEACR, by assessing motif enrichment—the degree to which identified peaks contain the expected transcription factor binding motifs. Our benchmarking highlights the impact of peak callers and preprocessing strategies on the analysis of CUT&RUN data. By evaluating the robustness and limitations of two widely-used peak callers, we provide practical guidance on tool selection and data preparation. We expect that our work will guide more informed choices in CUT&RUN analysis and support the development of improved computational methodologies.

## B-009: Machine Learning Analysis of Workplace Experiences AmongFrontline Healthcare Workers During COVID-19

**Authors List:**
Anupama Chalil Velluva, Thompson Rivers University, Canada
Amrita Ajay Sagar, Thompson Rivers University, Canada
Zeenathun Nusra Nazar, Thompson Rivers University, Canada
Michelle Smith, Interior Health, Canada, Canada
Bala Nikku, Thompson Rivers University, Canada
**Nisha Puthiyedth**, Thompson Rivers University, Canada

**Presentation Overview:**

The COVID-19 pandemic substantially altered working conditions for frontline healthcare workers (FHCWs), creating complex challenges related to workplace demands, resource access, and organizational support. This study examines patterns in workplace experiences among FHCWs in British Columbia using survey data from over 430 respondents across all five provincial health authorities. Relationships between demographic characteristics and reported access to COVID-19 training, testing, and vaccination, as well as workplace experiences, were evaluated using chi-square tests and logistic regression. Composite measures derived from Likert-scale survey items were used to summarize overall workplace strain. An interpretable machine learning approach was applied to identify factors associated with higher risk profiles. A Random Forest classifier achieved an accuracy of 69% (F1 = 0.68, AUC = 0.73) in distinguishing between outcome groups. Model interpretation using SHAP highlighted work–life balance, pandemic-related stress, COVID-19 worry, and perceived organizational support as influential contributors. Additional analyses indicated modestly higher outcome scores among FHCWs with experience in long-term care settings. Overall, this study demonstrates the value of combining traditional statistical analysis with explainable machine learning to better understand workplace dynamics in healthcare settings. The findings support data-driven approaches to workforce planning and organizational decision-making during public health emergencies.

**B-011: Similar Maturation Patterns of Gut Bacterial and PhageCommunities in infants in a region with high-stuntingprevalence.**

**Authors List:**
**Laura Carolina Valera**, Department of Microbiology and Immunology, McGill University, Montréal. Canada., Canada
Alejandro Reyes, Computational Biology and Microbial Ecology Group, Universidad de los Andes, Bogotá, Colombia., Colombia
Corinne Maurice, Department of Microbiology and Immunology, McGill University, Montréal. Canada, Canada

**Presentation Overview:**
Undernutrition remains a major cause of mortality among children under five, especially in low- and middle-income countries. Stunting, which affects nearly 23% of children globally, represents the most prevalent form of undernutrition and leads to poor physical and cognitive development, along with increased risk of developing chronic diseases in adulthood. To understand the microbial changes taking place in this condition, we investigated gut bacterial and phage community maturation during early life in a cohort of 219 children from a region with high prevalence of stunting. Our metagenomic analysis unveiled nuances previously unseen in more severe forms of undernutrition such as wasting. We observed that bacterial and phage communities of stunted and healthy children show similar composition and abundance over time at the genus and species levels, thus highlighting the need for higher resolution analysis. However, preliminary phage strain-level analyses reveal a trend towards reduced strain prevalence in stunted infants. These findings suggest subtle, yet critical, differences in eco-evolutionary dynamics that are only apparent with sub-species resolution. Our work validates a robust strategy for high-resolution viral analysis and highlights that identifying strain-specific phage patterns may be essential to uncover mechanistic links to stunting and guide targeted microbial interventions.

**B-013: Integrative computational–experimental pipeline uncoverspeptidase inhibitors against cryptococcal virulence andantifungal resistance**

**Authors List:**
**Jennifer Geddes-McAlister**, University of Guelph, Canada
Davier Gutierrez-Gongora, University of Guelph, Canada
Jacob Hambly, University of Guelph, Canada
Norris Chan, University of Guelph, Canada
Oscar Romero, University of Guelph, Canada
Michael Woods, University of Guelph, Canada
Esther Olabisi-Adeniyi, University of Guelph, Canada
Aimee Dawe, University of Guelph, Canada
Maria Mantilla, University of Montreal, Canada
Greg Wasney, The Hospital for Sick Children, Canada
Jared Deyarmin, Thermo Fisher Scientific, United States
Stephanie Samra, Thermo Fisher Scientific, United States
Pedro Valiente, University of Toronto, Canada
Melanie Alpaugh, University of Guelph, Canada
Adnane Sellam, University of Montreal, Canada
Andrew Hamilton-Wright, University of Guelph, Canada

**Presentation Overview:**
Fungal infections are a major global health challenge, with current antifungal therapies limited by toxicity, cost, and resistance. For Cryptococcus neoformans, key virulence factors that initiate and sustain infection are regulated by fungal peptidases to produce a polysaccharide capsule, promote immune evasion, and support antifungal resistance. These peptidases represent promising targets for antivirulent therapeutic strategies. Here, we developed a computational pipeline to predict and design peptide- and protein-based inhibitors against cryptococcal peptidases. Specifically, we targeted three virulence-associated peptidases: Rim13 (cysteine), May1 (aspartic), and CnMpr1 (metallo). Cysteine peptidase inhibition decreased capsule/cell size ratios without impeding fungal growth and reduced fungal survival within macrophages. Similarly, aspartic peptidase inhibition enhanced fungal clearance within alveolar macrophages and disrupted biofilm formation with additive effects towards fluconazole susceptibility in resistant strains. Additionally, metallopeptidase inhibition through catalytic zinc chelation and blocked substrate binding led to enhanced enzymatic inhibition and reduced in vitro blood-brain barrier crossing. Moreover, an in vivo larval model assessing inhibitor efficacy produced additive effects with fluconazole and lacked host cell cytotoxicity and fungicidal properties, reinforcing anti-virulence mechanisms and therapeutic potential while limiting the evolution of resistance. Further, global proteome profiling of inhibitor treated cells defined a mechanism of cell wall disruption, impeding fungal virulence. Taken together, the designed peptidase inhibitors exhibited potent antifungal activity without harming mammalian cells, establishing a predictive framework for rational scaffold design of next-generation antifungals that disarm the pathogen enabling immune-mediated clearance.

## B-015: Optimizing snoRNA Sequencing: Comparative Analysis of Reverse Transcriptases and Library Preparation Approaches

**Authors List:**
**Morgane Govone**, Université de Sherbrooke, Scott Group, Canada
Sonia Couture, RNomics Platform, Canada
Bergeron Danny, RNomics Platform, Canada
Sherif Abou Elela, Université de Sherbrooke, Laboratoire Abou Elela, Canada
Michelle S. Scott, Université de Sherbrooke, Scott Group, Canada

**Presentation Overview:**
High-throughput RNA sequencing (RNA-seq) is central for transcriptome profiling, yet detecting and quantifying structured non-coding RNAs remain challenging. Small nucleolar RNAs (snoRNAs), although highly abundant and essential for ribosome biogenesis, are systematically under-represented in standard ribodepleted short-read RNA-seq datasets. This bias is attributed to limitations introduced during library preparation and reverse transcription. In this study, we compare multiple reverse transcriptases (RTs) and library preparation strategies for snoRNA sequencing within the context of the entire ribodepleted transcriptome. We evaluate group II introns-derived enzymes (TGIRT, Induro), a retrotransposon-based RT (SEQUOIA), and optimized M-MLV mutants (SuperScript IV, Maxima, Qzyme, NEBNext), using matched RNA samples and standardized analysis pipelines. Global transcriptome analyses reveal RT-dependent biases affecting gene detection and quantification. Focusing on snoRNAs, we observe an enzyme-specific difference of expression profiles. These differences are not strongly explained by canonical snoRNA features, suggesting complex enzyme-linked biases. Preliminary coverage analyses indicate non-uniform read distributions along snoRNA sequence and the presence of putative transcript extensions, raising questions about technical artifacts and previously uncharacterized snoRNA forms. Overall, this work highlights the critical impact of reverse transcription and library preparation on snoRNA profiling and provides a methodological framework for improving the representation of structured non-coding RNAs in RNA-seq data.

**B-017: Transcriptomic and Genomic Profiling Reveals DistinctMolecular Features of Metastatic Myxoid Liposarcoma**

**Authors List:**
**Huilin Niu**, Department of Biochemistry, Western University, Canada
Alexandra Kollara, Lunenfeld-Tanenbaum Research Institute, Canada
Nalan Gokgoz, Lunenfeld-Tanenbaum Research Institute, Canada
Yael Babichev, Lunenfeld-Tanenbaum Research Institute, Canada
Brendan C. Dickson, Sinai Health System; Department of Laboratory Medicine and Pathobiology, University of Toronto, Canada
Jay S. Wunder, Sinai Health System; Division of Orthopaedic Surgery, University of Toronto, Canada
David G. Kirsch, Medical Biophysics, University of Toronto; Radiation Oncology, UofT; RadiationMedicine Program, Princess Margaret Cancer Centre, University Health Network, Canada
Irene L. Andrulis, Lunenfeld-Tanenbaum Research Institute; Department ofMolecular Genetics, University of Toronto, Canada
Rebecca Gladdy, Lunenfeld-Tanenbaum Research Institute; Laboratory Medicine and Pathobiology, University of Toronto; Sinai Health System, Canada
Kim M. Tsoi, Lunenfeld-Tanenbaum Research Institute; Sinai HealthSystem; Division

**Presentation Overview:**
Myxoid liposarcoma (MLPS) is a mesenchymal malignancy arising from adipose tissue driven by the FUS:CHOP. MLPS affects approximately 1 in 500,000 individuals per year. While the 5-year overall survival for localized disease is high (80–90%), survival drops to ~50% once metastasis occurs. We assembled and analyzed the first and largest cohort of untreated primary MLPS tumours profiled by whole-genome sequencing and bulk RNA sequencing. Comparative analyses between MLPS (n=46) and non-MLPS (n=44) adipose tumours found epigenetic dysregulation and metabolic reprogramming in MLPS. Despite similar mutational burdens, a distinct cluster enriched for metastatic primary tumours (n=11) showed upregulation of metabolic, proliferative, and cell-cycle pathway genes. The only non-metastatic tumours (n=4) within this cluster had follow-up times of less than 30 months, suggesting that this gene set may be predictive of future metastasis. In addition, we identified an MLPS subtype characterized by strong interferon signaling despite the low immune cell content of MLPS tumours. Genomic analyses confirmed recurrent alterations in known cancer drivers, including chromatin regulators, APC, and PIK3CA. Together, these findings identify epigenetic and metabolic dysregulation as central features of MLPS biology and highlight candidate biomarkers for metastatic risk stratification and potential therapeutic targeting.

## B-019: Machine Learning-Based Analysis Reveals Genetic Correlates of Isoform Diversity Across Eukaryotic Species

**Authors List:**
**Quinton Brohart**, University of Ottawa, Canada
Arvind Mer, University of Ottawa, Canada

**Presentation Overview:**
Alternative splicing (AS) influences transcriptomic and proteomic diversity in eukaryotes; however, the features that influence isoform repertoires remain incompletely understood. Here, associations between gene-level architectural, regulatory, and functional properties alongside isoform count in the human genome are systematically explored. We integrate exon-intron structure, untranslated region (UTR) length, GC and CpG content, and protein-protein interaction (PPI) data to identify conserved and lineage-specific patterns underlying isoform diversity, with comparative analyses performed with Mus musculus, Danio rerio, and Drosophila melanogaster genomic data. Across all, isoform-rich genes were characterized by expanded intron content and shorter exon content, indicating gene architecture is a primary determinant of isoform diversity. In contrast, regulatory features exhibited species- and biotype-specific associations. In humans, isoform count was positively associated with extended 3? UTR lengths and PPI connectivity but negatively associated with GC content and CpG island density, suggesting that regulatory flexibility and network integration contribute to, and are influenced by, isoform diversification while sequence composition imposes constraints. Similar but weaker trends were observed in mouse data, whereas zebrafish and Drosophila displayed distinct regulatory patterns, reflecting evolutionary divergence and annotation depth. These results demonstrate that isoform diversity is a structured property shaped by conserved architectural features and variable regulatory influences.

## B-021: Spatial Transcriptomics Reveals Antigen-PresentingMeningeal B Cells in Multiple Sclerosis

**Authors List:**
**Boris Tchatchoua Ngassam**, Department of Biochemistry, University of Western Ontario, London, ON, Canada
Cameron Whalen, Department of Microbiology and Immunology, University of Western Ontario, London, ON, Canada
Steven Kerfoot, Department of Microbiology and Immunology, University of Western Ontario, London, ON, Canada
Tallulah Andrews, Department of Biochemistry, University of Western Ontario, London, ON, Canada

**Presentation Overview:**
Meningeal B cells are known to be involved in driving central nervous system inflammation in multiple sclerosis (MS), yet their BCR-independent role within the meninges is poorly understood. While conventional B–T cell interactions promote antibody production, meningeal B cells may operate through non-canonical mechanisms to promote T-cell activation independent of their antibody-secreting role. To characterize these non?canonical B?cell functions, we used high?resolution Visium HD spatial transcriptomics on spinal cord sections from a mouse model MS. We identified distinct B-cell states, naïve-like, antigen presenting, and plasmablasts, within meningeal B-cell clusters. Cluster-associated B-cells had increased expression of APC genes including MHC class II and co-stimulatory molecules CD40, CD86. Clusters colocalized with lesion location enriched in microglia. The cluster B cells were frequently adjacent to T-cells in MS meninges, suggesting a role in T-cell activation via antigen presentation. Immunofluorescence confirmed these B cells and CD4? T cells co-localized, and interact via MHC II and SLAM, consistent with functional B–T synapses. Future work will leverage spatial transcriptomics and immunofluorescence to characterize the effect of these B-cell clusters on MS-associated lesion formation, aggravation and/or resolution.

## B-023: Whole Genome Segregation Analysis of Restless Legs Syndrome in French-Canadian Families

**Authors List:**
**Andrew Roth**, McGill University, Canada
Fulya Akçimen, National Institute on Aging, National Institutes of Health, United States
Patrick Dion, McGill University, Canada
Guy Rouleau, McGill University, Canada

**Presentation Overview:**
Restless legs syndrome (RLS) is a common yet underdiagnosed sensorimotor disorder affecting ~10% of the elderly population of North American and European ancestry. It causes an urge to move the legs, particularly at night, disturbing sleep and impacting mental and physical health. Although the pathophysiology of RLS remains unclear, substantial evidence supports a genetic contribution. Among French-Canadian (FC) RLS patients, 77.1% report a positive family history, indicating a significant genetic contribution in this population. Despite reported potential risk loci, causal coding variants segregating in multiplex families remain unestablished. We therefore analyzed whole-genome sequencing data from 25 FC families to identify rare, disease-associated variants. We clinically assessed affected individuals before sequencing. We processed genomic data using GATK joint genotyping and ANNOVAR annotation to integrate functional and population frequency data. We filtered for rare, deleterious protein-coding variants, identified segregating variants within families, and compared them across families and with known RLS risk loci. We identified 87 unique variants that segregate within one family and appeared in at least one individual from another family. None of these variants are pathogenic in clinical databases. This study establishes a family-based variant identification framework for RLS to illuminate its biological mechanisms and guide future neurobiological research.

**B-025: Single-cell analysis of the mechanisms underlying leukemictransformation in myeloproliferative neoplasms?**

**Authors List:**
**Wenqi Yang**, Department of Biochemistry, Schulich School of Medicine & Dentistry, Western University, Canada
Tallulah Andrews, Department of Biochemistry, Schulich School of Medicine & Dentistry; Department of Computer Science, Western University, Canada
Jenny Ho, Department of Medicine, Schulich School of Medicine & Dentistry, Western University, Canada

**Presentation Overview:**
Myeloproliferative neoplasms (MPNs) are at high risk of progression to acute myeloid leukemia (AML), occurring in up to 20% of patients with myelofibrosis. AML arising from this leukemic transformation (LT) have a median survival of 6 months. To investigate early changes preceding LT, we collected CD34? blood cells from 12 MPN patients, including 6 who later underwent LT and 6 who remained stable. We used single-cell transcriptomic and chromatin accessibility profiling were to define altered cell states and gene regulatory changes. We identified 12 major hematopoietic stem and progenitor cell populations, with an expansion of erythroid progenitors in LT patients. Among all cell populations, hematopoietic stem cells (HSCs) showed the strongest transcriptional and regulatory changes between LT and control samples. HSCs exhibited increased inflammatory signaling, including TNF-? and IL-2 pathways, along with altered transcription factor activity of BACH1 and AP-1 family members. These findings suggest that early regulatory disruption in HSCs may play a central role in LT and could serve as a biomarker to identify patients at increased risk. Future work will focus on validating key pathways as potential targets for preventative therapy.

## B-027: Computational docking analysis of dCRISPR-Cas12a regulatorycomponents targeting virulence-associated operons inPseudomonas syringae

**Authors List:**
**Jose Luis Silva**, Universidad de Guadalajara, Mexico
Ernesto Borrayo, Universidad de Guadalajra, Mexico
Bruno Alexis Velasquez, Universidad de Guadalajara, Mexico

**Presentation Overview:**
The regulation of bacterial virulence represents a promising alternative to traditional antimicrobial strategies, particularly in phytopathogenic systems where selective pressure leads to rapid resistance. As a model, Pseudomonas syringae employs complex regulatory networks to control virulence associated operons, such the type III secretion system (T3SS) and toxin production. Specifically, we employed computational docking to assess the efficiency of dCas12a-mediated silencing of key regulatory targets: the T3SS master regulator hrpL; and the virulence effector cfl to establish their suppression as predictive and durable disease management. Guide RNA's were designed targeting 23-nt protospacer sequences flanking TTTV-PAM sites. R-loop structural models were generated through homology-based modeling using Cas12a crystallographic templates (PDB: 5XUS) with sequence-specific mutations introduced via PyMOL. Molecular docking assessed protein-nucleic acid interactions (HDOCK server). Thermodynamic specificity was validated through RNA-DNA hybridization analysis (ViennaRNA/RNAcofold). Docking simulations yielded favorable binding scores ( -340 kcal/mol, confidence > 0.97) for both targets, with structural models confirming R-loop geometric compatibility. Hybridization analysis demonstrated robust sequence discrimination: cognate targets showed strong hybridization (?G: hrpL -40.5, cfl -50.6 kcal/mol) versus scrambled controls (?G ? -12 kcal/mol), yielding ??G > 28 kcal/mol. This computational framework establishes foundation for network-level dynamic modeling and experimental validation toward predictive, non-lethal pathogen control

## B-029: Characterizing Highly Conserved Fragments in 3'UTRs via Statistical and Transfer Learning Approaches

**Authors List:**
**Eric Ho**, Lafayette College, United States

**Presentation Overview:**
3' untranslated regions (3' UTRs) serve as regulatory platforms that modulate translation, mRNA localization, and stability through the binding of regulators - RNA-binding proteins and miRNAs. Their binding sites are often identified through orthologous regions among species. A separate but related discovery is the ultraconserved elements (UCEs) detected in human, rat, and mouse genomes two decades ago. However, knowledge about their functions is limited. Perplexingly, alterations in UCEs in mouse embryos produced no observable phenotypic differences. The majority of UCEs are non-coding, though ~8% are located in the 3'UTRs. Given the importance of 3'UTRs in gene regulation, we use a computational approach to identify highly conserved fragments (CFs) that exhibit >50 bp and >90% identity in 3'UTRs across diverse mammals. CFs are not composed of simple repeats or low-complexity regions common to mammalian genomes. Using a transformer-based foundational genomic model, CFs are characterized as A/T-rich and distinguishable from the 3'UTR background. 36 human CFs from 25 genes are significantly depleted in variations, enriched in neuronal tissues, and play roles in neurodevelopment and RNA processing. Our findings expand on existing studies that attribute UCEs primarily to enhancer function, suggesting a new path to explore the biological roles of UCEs in 3'UTRs.

## B-031: Evaluating the influence of annotation and statistical frameworks on pathway analysis results

**Authors List:**
Ritu Verma, University of Guelph, Canada
Jeff Caswell, University of Guelph, Canada
Jeremy Simpson, University of Guelph, Canada
Anthony Mutsaers, University of Guelph, Canada
Sonja Fonfara, University of Guelph, Canada

**Presentation Overview:**

Hypertrophic cardiomyopathy (HCM) is a common acquired cardiac disease in cats and humans, yet its molecular mechanisms is not fully understood. RNA sequencing was performed on left ventricular and atrial tissue from healthy cats and cats with HCM, followed by pathway analysis using multiple tools, including DAVID, g:Profiler, Ingenuity Pathway Analysis (IPA), and Gene Set Enrichment Analysis (GSEA). Considerable variability was observed across pathway analysis approaches. Using default databases, DAVID and g:Profiler identified few mostly generalized biological processes, consistent with limited feline annotation. Using a custom feline background gene database for g:Profiler increased the number and diversity of detected pathways. IPA found numerous canonical pathways; filtering criteria influenced pathway rankings and statistical significance, and potential biologically relevant pathways were not always highly ranked. GSEA, using a custom feline pathway database, identified multiple biologically relevant Gene Ontology terms associated with cardiac structure, metabolism, and remodelling. Across methods, pathway detection and ranking were strongly influenced by database completeness, statistical framework, and filtering stringency, with increased thresholds relevantly reducing pathway detection. These results demonstrate the implication of database content and methodological choices when carrying out pathway analysis, which is of importance for transcriptomic studies in particular when analysing less-annotated genomes.

**B-033: Automatic characterization of regulatory elements in thehuman genome using multimodal integration of '-omics' data**

**Authors List:**
**Julien Raynal**, CNRS, France
Charles-Henri Lecellier, CNRS, France
Laurent Bréhélin, CNRS, France

**Presentation Overview:**
Deciphering the DNA cis-regulatory code—critical for gene expression regulation—remains a major challenge in genetics and cancer research. While deep learning and statistical models now predict gene regulation and expression from DNA sequences, they primarily rely on dominant signals often located at promoters and enhancers. Yet, key regulatory features (e.g., open chromatin, transcription factor binding sites, and disease-associated variants) also reside outside these regions, where experimental data are sparse or noisy. As a result, current models struggle to assess the functional impact of genetic variants in these understudied areas. Building on our previous work demonstrating the predictability of transcription beyond canonical regions [Grapotte et al., Nat. Comm 2021], we aim to move beyond epigenetic segmentation (e.g., enhancers/promoters) by developing automated functional annotation methods. These methods leverage machine learning models optimized for specific genomic contexts. Here, I present models predicting CAGE-based transcription in unannotated regions and highlight their divergence from promoter- or enhancer-trained models. I then introduce a clustering approach to group regions by shared features (using different embeddings) and assign each cluster an optimal predictive model. This strategy promises genome-wide variant effect assessment, addressing the limitation of models restricted to canonical regulatory regions.

## B-035: A Fragment Size Correction Algorithm for ImprovedNucleosome Positioning from MNase-Seq Data

**Authors List:**
**Anisur Rahman**, Dept. of Pathology and Molecular Medicine, Queen's University, Kingston, Canada, Canada
Shuxiang Li, Dept. of Pathology and Molecular Medicine, Queen's University, Kingston, Canada, Canada
Anna Panchenko, Dept. of Pathology and Molecular Medicine, Queen's University, Kingston, Canada, Canada

**Presentation Overview:**
In eukaryotic genomes, nucleosomes are arranged like ""beads on a string,"" with approximately 147 base pairs of DNA wrapped around histone proteins. The genomic distribution of nucleosomes, known as nucleosome positioning, is not random and is governed by diverse regulatory factors. However, these positions are dynamic and can vary between cells, making accurate identification challenging. Nucleosome positioning provides insight into chromatin architecture, gene regulation, and the epigenetic landscape of a cell population, which is crucial for understanding diseases like cancer and developing diagnostic and therapeutic solutions. Micrococcal nuclease (MNase) digestion is commonly used to isolate DNA fragments bound to nucleosomes, and various computational tools are employed to infer nucleosome positions from the resulting sequencing data. However, MNase exhibits sequence bias, and the inherent variability of nucleosome positions, combined with methodological constraints, introduces significant noise in the sequencing data. Although current methods apply different smoothing and filtering strategies to mitigate these issues, their accuracy remains limited. We propose NuFix, an algorithm that computationally adjusts fragment lengths by incorporating the distribution of surrounding fragments, significantly enhancing nucleosome positioning accuracy. After applying NuFix to MNase-seq data, we identified more nucleosomes in the human genome with better accuracy and confidence.

## B-037: Discovering Genetic Biomarkers of Neuronal VulnerabilityUsing Post-Mortem Brain Datasets

**Authors List:**
**Xiaolin Zhou**, University of Toronto; Centre for Addiction and Mental Health, Canada
Daniel Kiss, Centre for Addiction and Mental Health, Canada
Isabel Kerrebijn, University of Toronto; Lunenfeld-Tanenbaum Research Institute, Mont Sinai, Canada
Vanessa Gonçalves, University of Toronto; Centre for Addiction and Mental Health, Canada
Yuliya Nikolova, University of Toronto; Centre for Addiction and Mental Health, Canada
Shreejoy Tripathy, University of Toronto, Canada

**Presentation Overview:**
Background and Objectives Neurodegenerative diseases, including Alzheimer's disease (AD), exhibit selective vulnerability of specific neuronal subtypes. Somatostatin-expressing (SST) interneurons are particularly affected, yet the genetic determinants of inter-individual variability in brain cell type proportions (CTPs) remain poorly understood. This study aims to identify common genetic variants underlying variability in human brain CTPs (Aim 1), and evaluate how these variants relate to neuropsychiatric traits (Aim 2). We hypothesize that neuronal and non-neuronal subtypes exhibit distinct genetic susceptibilities. Methodology For Aim 1, bulk RNA sequencing data from the ROSMAP post-mortem brain cohort (n = 1,168) were used to estimate CTPs using validated computational deconvolution. CTPs were analyzed as quantitative traits in genome-wide association studies (GWAS) and meta-analyzed across independent post-mortem brain cohorts (total n > 3,500), accounting for sex as a biological variable. Regarding Aim 2, associated loci were prioritized through fine-mapping, regulatory colocalization, and integration into polygenic risk scores (PRS). Results Meta-analysis identified 66 single-nucleotide polymorphisms associated with SST interneuron proportion variability at a suggestive threshold (p 1×10??), including loci within genes implicated in neuronal maintenance. Significance These findings demonstrate that common genetic variation contributes to cell type–specific neuronal vulnerability and support development of predictive biomarkers for precision neuropsychiatric research.

## B-039: Evaluating the Performance of Peptide-HLA BindingPredictors Reveals Gaps in Current Tools

**Authors List:**
**Issam Saadie**, Université de Montréal, Canada
David Hamelin, Université de Montréal, Canada
Jean-Cristophe Grenier, Montreal Heart Institute, Canada
Julie Hussin, Université de Montréal, Canada
Raphaël Poujol, Montreal Heart Institute, Canada

**Presentation Overview:**
Computational tools for predicting immune responses have become indispensable to immunological research, an impact made even more evident in the wake of the COVID-19 pandemic. Among those tools are those that predict the binding that occurs between a pathogen's peptide, called an epitope, and the human leukocyte antigen (HLA), getting us a step closer to detecting cytotoxic T-cell's immune response to virus infections. We aim to use these tools to predict immunogenic SARS-CoV-2 epitopes. Toward that goal, we evaluate the performance of three widely used peptide-HLA binding predictors: NetMHCpan, MixMHCpred and MHCFlurry. We conducted a series of comprehensive benchmarks using data from the immune epitope database (IEDB), giving us the current state of the art dataset to select the best performing tools. These benchmarks uncover a major weakness in the tools' ability to predict novel SARS-CoV-2 epitopes and show variable performance depending on epitope origin, viral- or cancer-based. We conclude by underscoring the need for more specialized tools tailored to viral epitopes and for more diverse, better-annotated immunological datasets to support future development.

## B-041: Skin Lesion Diagnosis using Deep Learning

**Authors List:**
**Penelope DeFreitas**, Toronto Metropolitan University, Canada
Youcef Derbal, Toronto Metropolitan University, Canada

**Presentation Overview:**
The incidence of skin cancer is increasing across North America, with melanoma remaining one of the most aggressive and life-threatening forms. This study reviewed multiple deep learning approaches and identified ResNet-50 as a comparatively weaker performing baseline for skin cancer classification. Experiments using the ResNet-50 architecture examined the impact of class balancing strategies, including targeted data augmentation, downsampling, SMOTE-Tomek, class weighting, and hybrid balancing, combined with fine-tuning ImageNet pretrained models. In addition to the baseline model, a ResNet-50 variant incorporating hierarchical diagnosis and a long short-term memory–based attention mechanism was evaluated. The ISIC 2017 dataset was used for Phase 1 hyperparameter tuning, followed by Phase 2 comparative class balancing and model experimentation using the ISIC 2018 dataset. Model performance was assessed using key metrics, including accuracy, sensitivity, and AUC-ROC. Results showed that the ResNet-50 variant with hybrid class balancing achieved the most consistent performance, particularly improving the detection of minority lesion classes. These findings demonstrate the value of integrating targeted class balancing strategies, careful hyperparameter optimization, and advanced attention mechanisms to improve classification robustness. Furthermore, the inclusion of explainable artificial intelligence enhanced model interpretability, supporting the potential for real world clinical adoption of AI-assisted diagnostic systems.

## B-043: Benchmarking AlphaFold 3, AlphaFold-Multimer, and ColabFoldfor Heterodimeric Complex Prediction in Arabidopsis thaliana

**Authors List:**
**Yifan Wu**, Cell and Systems Biology, University of Toronto, Canada
Nicholas Provart, Cell and Systems Biology, University of Toronto, Canada

**Presentation Overview:**
AlphaFold has transformed biological research by enabling accurate modeling of protein structures and protein–protein interactions (PPIs). Despite its widespread application, its performance in plant PPI modeling remains underexplored. Following the release of AlphaFold 3 and its source code, we benchmarked three AlphaFold variants, AlphaFold 3 (AF3), AlphaFold-Multimer (AFM), and ColabFold (CF), using the model plant species Arabidopsis thaliana. In the absence of a perfectly matched benchmark dataset, we constructed a dataset of 201 heterodimeric interfaces from the Protein Data Bank (PDB). Overall, the three methods exhibited comparable success rates and mean accuracies. Notably, AF3 generated a higher proportion of both high-quality and incorrect models than AFM and CF. Although average performance was similar, individual benchmark job outcomes varied across programs, with 24 cases in which only one program produced a successful prediction. We further investigated the influence of template availability and intrinsic interface properties on model quality. In particular, binary interactions derived from large assemblies (“megacomplexes”) displayed distinct properties, and template-based modeling substantially improved their prediction accuracy. Finally, we evaluated confidence metrics and model selection strategies. Collectively, our results demonstrate AlphaFold’s utility for modeling Arabidopsis PPIs and provide practical insights into its broader applications.

## B-045: A novel homozygous variant in AHCY causes a rare musculardystrophy: A new case

**Authors List:**
**Valerie Triassi**, University of Montreal Hospital Research Center (CRCHUM), Montreal, Quebec, Canada, Canada
Benjamin Ellezam, Pathology, CHU Sainte-Justine, Department of Pathology and Cell Biology, Université de Montreal, Quebec, Canada, Canada
Khadije Jizi, Research center Azrieli CHU Sainte-Justine, University of Montreal, Montreal, Quebec, Canada, Canada
Sébastien Jacquemont, Research center Azrieli CHU Sainte-Justine, University of Montreal, Montreal, Quebec, Canada, Canada
Julie Dery, Research center Azrieli CHU Sainte-Justine, University of Montreal, Montreal, Quebec, Canada, Canada
Cam-Tu Emilie Nguyen, Neurology, Centre Hospitalier Universitaire Sainte-Justine, Montreal, Quebec, Canada., Canada
Martine Tetreault, University of Montreal Hospital Research Center (CRCHUM), Montreal, Quebec, Canada., Canada

**Presentation Overview:**
Muscular dystrophies are rare inherited disorders, and despite advances in diagnostic pipelines based on targeted panels and exome sequencing, many patients remain undiagnosed due to the genetic heterogeneity of these conditions and the highly polymorphic nature of associated genes. These approaches often fail to capture novel or poorly annotated genes, and interpretation of genetic variation remains challenging highlighting the need for personalized bioinformatic analyses. We present a bioinformatic workflow used to resolve a previously undiagnosed neuromuscular disease in a young female patient. After negative results from first-line genetic testing, transcriptomic data generated from the patient's muscle tissue were analyzed incorporating variant detection, splicing analysis, and in silico pathogenicity prediction. This approach identified a novel homozygous missense variant in the AHCY gene (c.131C>T; p.Pro44Leu). Multiple prediction tools supported its pathogenicity and evolutionary conservation, including GERP, PhastCons, SIFT, PolyPhen-2, CADD and MutationTaster. Transcriptomic data allowed exclusion of splicing effects associated with this variant. Integration of these results with previously published cases strengthened the variant's relevance and enabled access to a potential therapeutic protocol. This study demonstrates how personalized bioinformatic workflows can overcome limitations of conventional diagnostic pipelines, improve variant interpretation, and increase diagnostic yield in rare diseases.

## B-047: State Dependent Communication Between Muscle Stem Cells and Their Niche

**Authors List:**
**Melisa Acun**, University of Ottawa, Canada
Arvind Mer, University of Ottawa, Canada

**Presentation Overview:**
Skeletal muscle regeneration depends on coordinated communication between muscle stem cells (MuSCs) and their surrounding niche. How signaling differs between quiescent (QSC) and activated (ASC) MuSC states, and how it is remodeled in dystrophic muscle, remains incompletely understood. We performed an integrative analysis of cell-cell communication across multiple single-cell transcriptomic datasets from normal and dystrophin-deficient (mdx) mouse muscle. Using expression-supported ligand-receptor filtering, we quantified signaling programs and associated transcriptional responses. In normal muscle, ASC and QSC exhibited clearly separated communication profiles. ASCs preferentially engaged extracellular matrix and inflammatory pathways, including ANGPTL, CXCL, and EGF signaling, whereas QSCs were enriched for adhesion and regulatory programs such as NCAM and PROS. Among ASC-QSC signaling pathways, 67% were retained in normal muscle but lost in mdx, while 33% were preserved in mdx, indicating remodeling of the dystrophic niche. Notably, quiescent MuSC signaling in mdx more closely resembled activation-associated programs than normal quiescence, indicating reduced separation between MuSC states. Approximately 82% of mdx quiescent MuSCs displayed activation-associated transcriptional signatures compared with 23% in normal muscle (Fisher’s exact test p 2.2×10?¹?). Together, these results demonstrate that dystrophic conditions collapse quiescent MuSC identity toward activation-like states, revealing disease-associated niche signaling that may impair regenerative stability.

## B-049: Mapping genetic modifiers of human disease inpopulation-scale biobanks

**Authors List:**
**Marcia S. Jude**, Department of Molecular Genetics, University of Toronto, Canada
Daxin Wu, Department of Molecular Genetics, University of Toronto, Canada
Hamed Heydari, University of Toronto, Canada
Charles Boone, Department of Molecular Genetics, University of Toronto, Canada
Gary Bader, Department of Molecular Genetics, University of Toronto, Canada

**Presentation Overview:**
Genetic interactions, or epistasis, play a critical role in shaping complex disease risk and may explain the incomplete penetrance and heterogeneous phenotypic effects of disease-associated variants. However, systematic identification of gene-gene interactions in large human cohorts has been limited by substantial computational demands, low statistical power due to the large number of possible interactions, and confounding by population structure and polygenic background. To address these limitations, we apply a scalable linear mixed model framework for gene-based interaction testing to whole-genome sequencing and phenotypic data from approximately 500,000 UK Biobank participants, focusing on quantitative endophenotypes that capture biological processes underlying disease. This approach enables systematic detection of epistatic effects while accounting for polygenic background, covariates, and family structure. We anticipate identifying modifiers of established disease genes, pathway-level interactions that buffer or amplify genetic risk, and gene networks that contribute to diversity in complex trait outcomes. Replication in independent cohorts, including the All of Us Research Program, will ensure that findings are robust, biologically meaningful and generalizable across populations. By uncovering biologically interpretable genetic networks, this work will refine models of complex trait architecture and improve our understanding of how genetic background modulates disease risk.

## B-051: FetchQuest: Secure and confidential agentic search forlaboratory quality management

**Authors List:**
**Jochen Weile**, Ontario Institute for Cancer Research, Toronto, Canada, Canada
Omar Alawamry, Ontario Institute for Cancer Research, Toronto, Canada, Canada
Carolyn Ptak, Ontario Institute for Cancer Research, Toronto, Canada, Canada
Melanie Courtot, Ontario Institute for Cancer Research, Toronto, Canada, Canada

**Presentation Overview:**
Clinical laboratories maintain thousands of interlinked quality management documents under strict regulatory frameworks, making rapid information retrieval daunting. Existing AI solutions are ineligible for these settings, as they rely on cloud-hosted proprietary large language models (LLMs) potentially exposing confidential data to third parties. FetchQuest is a local, secure, agentic search tool utilizing open-weights LLMs and high-quality vector indexing. The system offers two complementary modes: document search and question answering. The document search integrates semantic vector search, traditional keyword matching, PageRank, and customizable document importance scores, supporting both local repositories and SharePoint servers. To maintain accuracy with specialized terminology, FetchQuest features automatic acronym recognition and expansion. Evaluation demonstrated that manually curated correct answers appeared within the top five ranked results for 92.8% of search queries. The question-answering mode deploys a local LLM agent empowered by the document search. A neural sentry classifier screens inputs to prevent prompt-injection attacks, while automatic post-processing provides citations linking to original policy documents. When evaluated against a catalogue of audit questions, FetchQuest refused to answer 20.7% for lack of confidence, but correctly answered 91.3% of the remainder. FetchQuest provides a safe, confidential, and reliable information-retrieval service specifically designed for clinical laboratory documentation environments.

## B-053: De novo variant analysis of childhood-onset obsessive-compulsive disorder in the French-Canadian population

**Authors List:**
**Kate Bornais**, McGill University, Canada
Jay P. Ross, McGill University, Canada
Zoe Schmilovich, McGill University, Canada
Miranda Medeiros, McGill University, Canada
Charles-Étienne Castonguay, McGill University, Canada
Dan Spiegelman, McGill University, Canada
Bernard Boileau, Le Centre Hospitalier Universitaire Sainte-Justine, Canada
Jean-Jacques Marier, Le Centre Hospitalier Universitaire Sainte-Justine, Canada
Ghislain Laurin, Le Centre Hospitalier Universitaire Sainte-Justine, Canada
Patrick A. Dion, Montréal Neurological Institute-Hospital, Canada
Guy A. Rouleau, Montréal Neurological Institute-Hospital, Canada

**Presentation Overview:**
Childhood-onset obsessive-compulsive disorder (OCD) is a neuropsychiatric disorder with a strong genetic component. De novo variants (DNVs) have been shown to have a role in childhood-onset OCD, yet no DNV analysis has been performed in patients from a genetically isolated population. Here, we analyzed whole-exome sequencing data from 36 French-Canadian trios (including 36 probands and 72 unaffected parents) to identify de novo single nucleotide variants (dnSNVs) contributing to OCD risk. Using rigorous bioinformatics pipelines for variant calling, filtering, and annotation, we identified 34 dnSNVs harboured in 34 different genes. Four of these genes were previously associated with OCD, replicating their contribution to its risk. We also observed complete overlap between our 34 candidate genes and genes associated with 11 related neuropsychiatric disorders, supporting a shared underlying genetic susceptibility across psychopathologies. Among genes harbouring DNVs across three childhood-onset OCD cohorts, functional enrichment analyses identified an overrepresentation of genes involved in clathrin-dependent endocytosis (GO:0072583; p-adj = 0.0498) and phosphatidylinositol binding (GO:0035091; p-adj = 0.0431), offering potential biological mechanisms underlying childhood-onset OCD. Altogether, this study offers a framework for performing DNV analyses of complex disorders in genetically isolated populations and provides the first list of candidate childhood-onset OCD genes in the French-Canadian population.

## B-055: Benchmarking Analysis and Dashboard to Find an OptimalSequencing Strategy for Detection of Germline GeneticVariants in a Human Sample

**Authors List:**
**Jose Hector Galvez**, McGill University / Canadian Centre for Computational Genomics, Canada
Robert J. M. Eveleigh, McGill University / Canadian Centre for Computational Genomics, Canada
Soofia Garmeh, McGill University / Canadian Centre for Computational Genomics, Canada
Sarah J. Reiling, McGill University Health Centre, Canada
Mathieu Bourgey, McGill University / DNA to RNA (D2R), Canada
Ioannis Ragoussis, McGill University, Canada
Guillaume Bourque, McGill University / Canadian Centre for Computational Genomics, Canada

**Presentation Overview:**
From the advent of tandem parallel short-read (SRS) to more recent long-read sequencing (LRS) technologies, there is a wide diversity of sequencing options with unique strengths and weaknesses. Bioinformatics analysis pipelines and depth of coverage add layers of complexity with impacts on results and reproducibility. We have performed extensive benchmarking efforts to determine the best sequencing and analysis approach depending on the research objective. We used a Genome-in-a-Bottle (GIAB) sample to generate datasets with several commercial sequencing solutions, then analysed them using different pipelines. We evaluated each platform's ability to detect single nucleotide variants (SNVs), indels, and structural variants (SVs) by comparing the outputs to the GIAB truth-set and calculating the precision, recall, and F1 scores. Our results indicate that SRS platforms analyzed with the DRAGEN pipeline offer the highest accuracy for SNVs and indels in low and moderately complex genomic regions. Whereas, LRS technologies produce the best results for detecting SVs. We built a dashboard to facilitate exploration of our results through interactive filtering and visualization; it will also allow users to upload their own benchmark results to compare with ours. Taken together, these efforts are a valuable resource for groups deciding which technologies suit their research objectives best.

## B-057: CanDataMine: An R Programming Resource and Database forData Mining and Drug Discovery

**Authors List:**
**Samantha Majoros**, Ontario Institute for Cancer Research, Canada
Mitchell Shiell, Ontario Institute for Cancer Research, Canada
Joe Wang, Ontario Institute for Cancer Research, Canada
Justin Richardsson, Ontario Institute for Cancer Research, Canada
Richard Marcellus, Ontario Institute for Cancer Research, Canada
David Uehling, Ontario Institute for Cancer Research, Canada
Rima Al-Awar, Ontario Institute for Cancer Research, Canada
Shraddha Pai, Ontario Institute for Cancer Research, Canada
Melanie Courtot, Ontario Institute for Cancer Research, Canada
Lincoln Stein, Ontario Institute for Cancer Research, Canada
Quang Trinh, Ontario Institute for Cancer Research, Canada
Vida Talebian, Ontario Institute for Cancer Research, Canada

**Presentation Overview:**
Databases such as The Cancer Genome Atlas Program and DepMap provide large-scale genome sequencing data from patient samples and cancer cell lines. Bioinformatics tools are needed to integrate this data and mine for potential therapeutic targets. We developed CanDataMine, which includes a series of generalized and flexible R scripts that, when given a gene list, can format and integrate data across sources (cBioPortal, etc.), run analyses (differential expression, etc.) and output visualizations and tabular results. Documentation and workflow examples along with these scripts will be shared in a public GitHub repository. Furthermore, preprocessed data will be made available through a data portal so users can search for genes of interest. We preprocessed data (mutations, etc.) for 32 cancer types and developed an internal data portal with a web interface using the Overture platform (https://www.overture.bio/). Future development will integrate an AI-powered conversational interface enabling researchers without a programming background to discover data, connect with external resources, and generate insights via reproducible workflows. CanDataMine can serve as an open-source platform that not only allows users to uniformly analyze large amounts of public data but also improve the efficiency of generating new biological knowledge and the overall drug discovery process.

## B-059: Beyond Structure: Improving the Prioritization andInterpretation of Genetic Variants Using Protein Motion

**Authors List:**
**Nour Hanafi**, Department of Medical Biophysics, University of Toronto, Canada
Sushant Kumar, Department of Medical Biophysics, University of Toronto, Canada

**Presentation Overview:**
Monumental advances in genome sequencing have established it as a mainstay of clinical and biological research. However, this influx of data has created a bottleneck in understanding variant impact, as most sequenced variants remain uncharacterized. One approach to improving variant prioritization and impact is to contextualize the variant within the three-dimensional protein structure. This has been used to great success with in-silico predictors, as well as improved calculations of genetic constraint. However, these tools are optimized to identify regions critical to the static protein structure, and do not consider the dynamic nature of a protein. We have built upon this by capturing critical protein subregions related to protein motion. To do this, we have generated conformational ensembles for 11,438 human proteins using an AlphaFold2 flow-matching derivative, AlphaFlow. We used these ensembles to quantify short-range contact frequencies, long-range networks via perturbation response scanning, and conformational switch residues using angular distributions. Our results show that incorporating local and distant motion into genetic constraint calculations results in pathogenic variant prioritization that is improved over calculations that only use single structure. This can be used to improve diagnostic yield, and provide insight into the mechanistic underpinnings of rare diseases and cancer drivers.

## B-061: Multi-Objective Reinforcement Learning for De NovoMolecular Design Targeting Pin1

**Authors List:**
**Yan Yi Li**, Western University, Canada
Pingzhao Hu, Western University, Canada

**Presentation Overview:**
The prolyl isomerase Pin1 is a validated regulatory protein implicated in multiple cancer types, yet remains an underexplored therapeutic target. Machine learning-driven molecular design provides a scalable strategy to navigate the vast chemical space beyond conventional screening approaches. We present a multi-objective reinforcement learning framework for de novo drug-like molecule generation targeting Pin1. Building upon FREED++, a statistical machine learning model for molecular design, we integrate curated fragment libraries derived from FDA-approved drugs with adaptive multi-objective optimization to simultaneously optimize predicted target affinity, drug-likeness, and synthetic accessibility. Our framework combines cheminformatics constraints with policy gradient-based reinforcement learning to efficiently explore high-dimensional chemical space while balancing exploration and exploitation. Generated molecules were evaluated across five criteria: structural uniqueness, chemical diversity, pharmacochemical property filters, molecular docking scores against Pin1, and synthetic accessibility. Comparative benchmarking against baseline generative models, together with ablation analyses and molecular dynamics simulations, demonstrates improved generation of structurally diverse, synthetically feasible candidates with favorable predicted binding profiles. This work advances statistical machine learning approaches for therapeutic discovery and highlights the potential of reinforcement learning-guided molecular generation for scalable cancer drug discovery and design.

## B-063: In Silico Reconstruction of Head and Neck Cancer Tumour Microenvironment Through Single Cell Dynamics

**Authors List:**
**Sameena Karsan**, McGill University, Canada
Jun Ding, McGill University, Canada
Nicole Li-Jessen, McGill University, Canada

**Presentation Overview:**
Most existing cancer models treat tumours as uniform masses, yet patients possess unique tumour microenvironments (TMEs). This biological heterogeneity drives clinical outcomes, necessitating a deeper understanding of individual cell behaviors. In HPV+ head and neck cancers, cancer associated fibroblasts (CAFs) are postulated to shape the complex TME and facilitate immune exclusion. To investigate these dynamic interactions, we integrated an individual patient's single cell spatial transcriptomics with agent-based modeling (ABM). Spatial transcriptomics characterizes gene expression profiles and spatial context to reveal cellular identities and neighborhood organization. We then applied biologically informed deep learning to infer interaction probabilities and communication pathways from the single cell data. These pathways inform rules to guide the ABM, where each cell acts as an autonomous agent. Our framework dynamically predicted gene expression according to cell interactions, achieving a median cell-cell correlation of 0.56 between timepoints. Additionally, modelling tumour epithelial proliferation showed a significant correlation with the pseudotime trajectory of proliferative genes ($r = 0.93$, $p$ 0.001). This novel ABM framework creates a dynamic digital, patient-specific head and neck tumour representation. Validated using pseudotime trajectories, our ABM simulates complex intercellular and cell-TME interactions that could help formulate clinically relevant hypotheses regarding therapeutic response and resistance.

## B-065: Mitochondrial RNA Degradation Regulates Differentiation,Stemness, and Immune Sensitivity in AML

**Authors List:**
Neil MacLean, UHN, Canada
Jong Bok Lee, University of Calgary, Canada
Li Zhang, UHN, Canada
Mark D Minden, UHN, Canada
Vito Spadavecchio, Interlinked Therapeutics LLC, USA
Andrea Arruda, UHN, Canada
Xiao Ming, UHN, Canada
Li Ma, UHN, Canada
Dakai Ling, UHN, Canada
Yulia Jitkova, UHN, Canada
**Veronique Voisin**, UHN, Canada
Yongran Yan, UHN, Canada
Marcela Gronda, UHN, Canada
Ali Chegini, UHN, Canada
Karen Kai-Lin Fang, UHN, Canada
Rose Hurren, UHN, Canada
Kazem Nouri, UHN, Canada
Aaron D Schimmer, UHN, Canada
Geethu Emily Thomas, UHN, Canada

**Presentation Overview:**
Mitochondria play a central role in metabolism and are particularly important in acute myeloid leukemia (AML). Although the mitochondrial proteome comprises approximately 1,000–1,500 proteins, only 13 are encoded by the mitochondrial genome, with the remainder encoded in the nucleus and imported into the organelle. Bioinformatics analyses of large public transcriptomic datasets revealed a global increase in mitochondrial gene expression in AML compared with normal hematopoietic cells. However, comparing RNA-sequencing strategies uncovered a striking discrepancy for the 13 mitochondrially encoded genes comparing AML to normal: datasets generated using poly(A) enrichment showed reduced mitochondrial transcript abundance, whereas ribosomal RNA depletion datasets showed increased expression. Additional computational analyses and wet lab validation indicated that, in AML mitochondria, poly(A) tails promote RNA degradation rather than stabilization, resembling bacterial systems. Genes involved in mitochondrial RNA transcription, processing, and degradation were upregulated in AML. The SUV3/PNPase degradosome degrades poly(A)-mtRNA and double strand RNA (dsRNA) from bidirectional mitochondrial transcription. dsRNA?sequencing confirmed dsRNA accumulation and Nanostring profiling pinpointed activated interferon and immune pathways. Together, by intertwining bioinformatics and wet lab validations, we highlight altered mitochondrial RNA turnover in AML and identified a novel mechanism by which mtRNA metabolism regulates leukemia stem cell function and immune sensitization.

## B-067: The Pan-Canadian Genome Library: Creating resources to promote genomic research in Canada

**Authors List:**
**Alexandra Smith**, Ontario Institute for Cancer Research, Canada
Amirah-Iman Hicks, Pan-Canadian Genome Library, Canada
Kimberly Begley, Pan-Canadian Genome Library, Canada
Karen Cranston, University Health Network, Canada
Wyeth Wasserman, University of British Columbia, Canada
David Wishart, University of Alberta, Canada
Guillaume Bourque, McGill University, Canada
Ma'n H. Zawati, McGill University, Canada
Michelle D. Brazas, Ontario Institute for Cancer Research, Canada
The Pan-Canadian Genome Library, Pan-Canadian Genome Library, Canada
Julia Ortiz, McGill University, Canada

**Presentation Overview:**
Researchers in Canada advance our understanding of health and genetics through large-scale projects that generate vast amounts of genomic data. However, these researchers often face challenges ensuring the long-term sustainability and discoverability of this data. The Pan-Canadian Genome Library (PCGL) is a national data repository created to meet this need: the PCGL unifies, archives and facilitates the sharing of genomic data, with a commitment to international standards and equity in genomic research. PCGL also provides training opportunities to support the development of Canada's genomic workforce capacity. PCGL's Training & Outreach team aims to provide training for all PCGL audiences, including researchers, trainees, research participants, and the public. We work with these audiences to identify knowledge and skills essential to them, and catalog relevant existing resources. Our researcher training catalog covers topics from fundamental skills, such as data management and ethical research practices, to downstream-analyses for population genetics and precision healthcare. In collaboration with the Canadian Bioinformatics Hub, we develop bespoke training materials and offerings for PCGL users. To ensure our training promotes Inclusion, Diversity, Equity, and Accessibility, we work with the PCGL Data Diversity and Inclusion team, the Indigenous Trainee Circle for Bioinformatics Training, and partners in the research community.

## B-069: Gene programs driving inter- and intra-tumor heterogeneity in renal cell carcinomas

**Authors List:**
**Ariel Madrigal Aguirre**, McGill University, Canada
Hamed S. Najafabadi, McGill University, Canada
Minjun Kim, McGill University, Canada
Zohreh Mehrjoo, McGill University, Canada
Adrien Osakwe, McGill University, Canada
Elham Moslemi, McGill University, Canada
Matthew Dankner, McGill University, Canada
Yu Chang Wang, McGill University, Canada
Haig Djambazian, McGill University, Canada
Jonathan Spicer, McGill University, Canada
Fadi Brimo, McGill University, Canada
Peter Siegel, McGill University, Canada
Morag Park, McGill University, Canada
Jiannis Ragoussis, McGill University, Canada
Simon Tanguay, McGill University, Canada
Yasser Riazalhosseini, McGill University, Canada

**Presentation Overview:**
Renal cell carcinoma (RCC) displays extensive intra-tumor heterogeneity. Gene programs, defined as sets of co-expressed genes under shared regulatory influences, provide a powerful lens to interpret cellular diversity. Gene program discovery from single-cell data, however, is often compounded by technical noise and complex inter- and intra-sample sources of variation, especially in tumors. These obstacles have limited our understanding of the gene programs underlying RCC heterogeneity, their role in metastasis and their clinical significance. To address this, we generated a cellular map of transcriptional heterogeneity in RCCs, comprising over 85,000 single-cell expression profiles from twelve primary and nine metastatic patient tumor tissues, along with five patient-derived xenograft models. To identify gene programs, we developed a generative modeling framework based on convex non-negative matrix factorization. Applying this framework to malignant cells revealed functional submodules within canonical pathways, including Hypoxia and Epithelial-to-Mesenchymal Transition (EMT). Comparisons between primary and metastatic tumors uncovered increased activity of a complete EMT program and reduced proximal tubule identity in metastatic cells. Spatial analyses revealed that activation and loss of these metastasis-associated gene programs emerged early in disseminating malignant cells at the invasive tumor front. Overall, our cellular atlas highlights multiple gene programs associated with metastasis and prognosis.

**B-071: Breaking the Language Barrier: A Regional Model forBilingual Bioinformatics Training and Community Engagementin Québec**

**Authors List:**
**Canelle Schuhler-Husson**, Canadian Bioinformatics Hub/Université de Montréal, Canada

**Presentation Overview:**
The expansion of multi-omics data has made biology into a data-driven science, making training and collaboration in bioinformatics essential. In Canada, regional language barriers often limit the reach of shared learning resources, presented almost exclusively in English. To engage with the significant yet Francophone research community, it is imperative to strengthen collaboration and champion bilingualism as a driver of innovation and national cohesion. The Canadian Bioinformatics Hub and the Canadian Bioinformatics Workshops (CBW) have implemented a regional strategy in Quebec. In 2025, this approach focused on organizing hands-on workshops tailored to the regional bioinformatics research landscape. Notably, the first ever CBW French-language workshop was launched in collaboration with l'Institut de Recherches Cliniques de Montréal. In 2026, the development of community training events in French or with bilingual support, combined with the participation of French-speaking ECRs and trainees, will remove barriers for the Francophone research community and increase their engagement. By collaborating with regional partners, events will continue to reach a larger, more diverse audience. Establishing and growing a regional coordination and bilingual outreach is an essential strategy for building sustainable bioinformatics communities. This framework of expansion will provide a model strategy that could be adapted to languages other than French.

**B-073: Integration of human and mouse single-cell transcriptomes of the developing cerebellum nominates cells-of-origin for Group 3 and 4 medulloblastoma**

**Authors List:**
Ian Cheong, University of Toronto, Canada
Leo Lau, University of Waterloo, Canada
**Shraddha Pai**, University of Toronto, Canada

**Presentation Overview:**
The cerebellar rhombic lip neurogenic niche is critical for glutamatergic neurogenesis, and dysregulated differentiation of rhombic lip lineage cells is hypothesized to cause medulloblastoma (MB), a malignant pediatric cancer lacking targeted therapies. Humans have structural compartmentalization of the rhombic lip into a subventricular zone not seen in mice, and the common Group 4 subtype of MB is hypothesized to arise by dysregulation of EOMES+ unipolar brush cells (UBC) generated by this compartment. However, it is unknown whether humans have unique UBC populations not seen in the mouse and, if so, whether Group 4 MB tumour cells resemble these UBC populations. We integrated 336,598 human and mouse single-cell transcriptomes of the developing cerebellum (11–21 post-conception weeks in humans and E10–P14 in mice) and identified two subpopulations of UBC that are enriched in the human samples, relative to mice. One of these populations predominates at 11 post-conception weeks, upregulates RBFOX1 and FOXP2, is predicted to be driven by OTX2, SOX4, and SOX11, and is enriched for axonogenesis and neurodifferentiation pathways. We analyzed 27,735 single-cell transcriptomes from eight MB tumours and recapitulated the observation that Group 4 MB cells best resemble UBC. We then found that two-thirds of Group 3 and Group 4 MB UBC-like cells best resemble human-enriched RBFOX1-rich cells. Gene regulatory network analysis revealed that top regulators of Group 4 MB tumour cells include EOMES, SOX4, and SOX11. Our work provides initial evidence for human-enriched UBC states and suggests that SOX4 and SOX11 may drive neurogenesis in UBC and gene expression in a subset of Group 4 MB tumour cells. Our findings shed light on genetic determinants of human cerebellar expansion and suggest that human models may be needed to recapitulate the oncogenesis of Group 4 MB.

## B-075: CoreNet – coreness-aware node embedding to improve functional module discovery

**Authors List:**
**Abeera Fatima**, Charite university Berlin, Germany
Ralf Herwig, Max Planck Institute for Molecular Genetics, Germany

**Presentation Overview:**
Identifying functional context from sparse genetic signals is still a major problem in network biology. Node embedding methods map network nodes onto a vector space by preserving neighborhood structure through training on biased random walks using a skip-gram model for the prediction of node neighborhoods. However, conventional random walk strategies do not take into account higher-order network topology, such as node cores, limiting their ability to prioritize influential and biologically meaningful paths through the network. We present a novel, coreness-aware embedding framework that integrates graph topology into the process of generating random walks. Node coreness informs about whether a gene/protein belongs to dense network regions or rather the periphery, and we generate core-biased random walk dictionaries to preserve node neighborhoods. We applied our method to GWAS-trait genes and learned trait-specific embeddings using skip-gram optimization. We compare several similarity measures on the embedding space that judge for each gene its trait-specificity and, finally, derive functional network modules. We show that this selection leads to interpretable and enriched associations of the originally-unrelated GWAS genes with respect to biological pathways and functions. Thus, our approach improves reconstruction of biologically meaningful subnetworks and provides a scalable approach for disease module discovery and gene prioritization.

## B-077: Variant calling from long-read whole genome sequencing ofmetastatic lung tissue affirms genetic clonal homogeneityachieved with RHAMM-deficiency

**Authors List:**
**Joseph Butler**, University of Western Ontario, Canada
Audrey Goddard, University of Western Ontario, Canada
Cornelia Tolg, University of Western Ontario, Canada
Eva Turley, University of Western Ontario, Canada
Kathleen Hill, University of Western Ontario, Canada

**Presentation Overview:**
Breast cancer (BrCa) displays high genetic heterogeneity with intra- and inter-tumor clonal diversity, a characteristic prominent across subtypes contributing to aggressiveness, treatment resistance and prognosis. In a MMTV-PyMT mouse model of luminal BrCa, Rhamm loss was associated with reduced clonal heterogeneity in primary tumors and lung metastases with recurrent, inter-animal mutations in metastatic tissue. These inter-animal recurrent mutations are consistent with clonal homogeneity and one or a few dominant clones. Clonal homogeneity can be exploited for targeted treatment and gene therapy. To resolve the full extent of the variants within the dominant clonal genotype associated with Rhamm loss, long-read sequencing for bulk tissue was conducted. DeepVariant analyses revealed a 1.6-fold higher mutation burden in Rhamm-/- compared to Rhamm+/+ metastatic lung tissues (average of 23,159 vs. 14,349 variants, respectively; p=0.002). Rhamm-/- samples had low inter-animal variation; variants exclusively shared across all Rhamm-/- mice totaled 4078 of which 83% were VAF?0.8 (mean VAF=0.94). Only 269 variants were uniquely shared across all Rhamm+/+ mice with wider VAF distributions (36% were major clones, VAF?0.8). These data confirm a role for RHAMM in promotion of genetic diversity and suggest BrCa cancer therapeutic approaches capitalizing on RHAMM deficiency.

**B-079: Assembly-Free Metagenomic Profiling Coupled with MetabolicNetwork Analysis for Soil Remediation.**

**Authors List:**
**Harbinder Kaur**, School of Computational & Integrative Sciences, Jawaharlal Nehru University, New Delhi, India
Namrata Joshia, School of Computational & Integrative Sciences, Jawaharlal Nehru University, New Delhi, India
Andrew Lynn, School of Computational & Integrative Sciences, Jawaharlal Nehru University, New Delhi, India

**Presentation Overview:**
Natural disasters and industrial activities, such as mining and manufacturing, can cause severe, often irreversible soil degradation. Healthy soils function as complex ecosystems in which geological, hydrological, and biological components are tightly coupled through microbial metabolic interactions. Effective remediation, therefore, requires a systems-level understanding of how community function is disrupted and restored. We developed a network-based, assembly-free framework for rapid functional and taxonomic profiling of soil metagenomic and metatranscriptomic data. The framework uses a custom reference database built from curated Ensembl transcript sequences. By linking custom database entries to a locally built functional lookup table, we enable direct linkage between taxonomic identity and gene-level function. A weighted functional abundance matrix is generated directly from read-level data. This bypasses the complexities of transcript reconstruction and provides a faster approach to challenges such as the fragmented nature of complex microbiome datasets. This approach was motivated by our earlier work on post–coal-mining soils. There, assembly-based methods consistently recovered very low functional activity. These functional profiles are mapped to a global metabolic reference network to assess ecosystem disturbance using metrics such as connectivity, modular integrity, and pathway erosion. The framework then highlights critical functional gaps in degraded soils by pinpointing where metabolic flow from environmental substrates to biomass production is constrained. To address these gaps, a Network Expansion algorithm predicts targeted microbial taxa combinations that act as functional “bridges” to restore metabolic reachability. We benchmark this framework using a published Auto-Brewery Syndrome (ABS) metagenomic dataset, which shows clear shifts in microbiome organization and ethanol-associated metabolic activity. This makes the dataset useful for evaluating sensitivity, functional recovery, and biological consistency. The approach is further validated using simulated metagenomic datasets of intact and degraded soil communities, demonstrating its potential to guide microbiome-based soil remediation strategies.

## B-081: Mining the Medical Record: AI-Based Curation of Electronic Medical Records for Cancer Risk Prediction

**Authors List:**
**Tony Yu**, Ontario Institute for Cancer Research, Canada
Julia Sobotka, Princess Margaret Cancer Centre, University Health Network, Canada
Sarah Ridd, Princess Margaret Cancer Centre, University Health Network, Canada
Jochen Weile, Ontario Institute for Cancer Research, Canada
Trevor Pugh, Princess Margaret Cancer Centre, University Health Network, Canada
Mélanie Courtot, Ontario Institute for Cancer Research, Canada

**Presentation Overview:**
Introduction: Valuable data in electronic medical records is often underutilized as a significant portion is stored as unstructured text. To meaningfully use this data, manual curation into standardized formats is necessary, a time-intensive process that is unfeasible at scale. As Artificial Intelligence (AI) and Large Language Models (LLMs) have shown promising information extraction capabilities for medical texts, we investigate their potential to advance scalable data curation for tailored risk stratification in oncology. AI-Based Curation: LLMs prompted to extract clinical entities from patient notes of the MedAlign dataset have yielded promising extraction accuracies with F1-scores averaging 83%, while significantly reducing time necessary for data curation compared to manual methods. Enabling clinical data integration for prediction of cancer-onset: Using data extracted and structured by the LLM-based pipeline, the clinical parameters will be integrated into machine learning models to predict cancer-onset, assessing their performance for risk stratification. Future Direction: The accuracy for risk modelling will be validated using real world clinical outcomes from the hereditary cancer patient cohort, CHARM (cfDNA in Hereditary And High-Risk Malignancies). Integrating additional data modalities from this cohort including blood-based cell-free DNA sequencing features will be explored for its potential to improve risk modelling in this scalable, automated proof-of-concept.

## B-083: Strengthening bioinformatics capacity in Alberta

**Authors List:**
**Robyn Woudstra**, Canadian Bioinformatics Hub, Canada
Mark Berjanskii, Canadian Bioinformatics Hub, Canada
David S. Wishart, University of Alberta, Canada

**Presentation Overview:**
The bioinformatics landscape in Alberta is critical to the national bioeconomy, but faces key challenges: an increased demand for skilled researchers, information siloing, and the underrepresentation of Indigenous peoples. Through targeted regional training programs and community initiatives, the Canadian Bioinformatics Hub (CBH) is addressing these challenges. Bioinformatics is an integral part of a robust life sciences ecosystem; however, bioinformatics innovation is often bottlenecked by gaps in training. CBH operates through six nodes, delivering hands-on workshops that address the specific needs of regions. In 2025, CBH Alberta hosted three workshops, with plans for four workshops in 2026. Because research institutions in Alberta are geographically dispersed, sharing knowledge and building community is difficult. The Alberta node of CBH facilitates community building and cross-institutional communication via networking events, seminar series, and partnerships with local bioinformatics organizations including Genome Alberta, BioNet, yegBUG, and Youreka. These partnerships also foster collaboration and knowledge exchange between academia and industry. The Indigenous Training Circle for Bioinformatics Trainees (ITCBT) is another important component of CBH. The ITCBT creates opportunities for community support and training programs to be shaped by Indigenous students themselves. Together, these CBH initiatives are strengthening bioinformatics capacity and cultivating vibrant life science communities across Alberta.

## B-085: Rethinking Biomedical Triple Extraction: An Open-Vocabulary Benchmark with LLM-Based Evaluation

**Authors List:**
**Jessica Xia**, University of Toronto, Canada
Hannes Röst, University of Toronto, Canada

**Presentation Overview:**
Semantic triples (subject-predicate-object) are essential for structured knowledge representation. However, automated triple extraction remains a challenge, with ongoing debate regarding whether domain-specific models or general-purpose large language models (LLMs) are best suited for the task. Although LLMs demonstrate superior semantic reasoning, existing biomedical triple extraction benchmarks like BioRED rely on a predefined vocabulary, often favoring smaller, domain-specific models. We hypothesize that this perceived advantage is an artifact of the restrictive design of current benchmarks, which reward pattern recognition over dynamic reasoning. Here, we introduce a benchmark of 346 triples which we manually extracted from 20 diverse PubMed abstracts without a predefined vocabulary. For LLM-based extraction, Gemini-3-Flash was few-shot prompted to generate triples consistent with our manual annotations. To evaluate these triples against the manual annotations, we re-employed Gemini-3-Flash to identify semantic matches, accommodating linguistic variability. In a direct comparison on our benchmark, LLM-based triple extraction (Gemini-3-Flash) outperformed a state-of-the-art domain-specific model (BioREx) (F1 = 0.682 vs. 0.317; adjusted p 0.001). Interestingly, our benchmark sample was large enough to identify statistically significant performance differences despite its moderate size. Our results suggest that few-shot LLMs can surpass domain-specific models when evaluated on open-vocabulary benchmarks for structured biomedical knowledge extraction.

**B-087: A Declarative Configuration-Driven Framework to HarmonizeResearch Data into the Pan-Canadian Genome Library (PCGL)**

**Authors List:**
**Linda Xiang**, Ontario Institute for Cancer Research, Canada
Edmund Su, Ontario Institute for Cancer Research, Toronto, Canada
Hardeep Nahal-Bose, Ontario Institute for Cancer Research, Toronto, Canada
Alexis Li, Ontario Institute for Cancer Research, Toronto, Canada
Daniel de Maria, Ontario Institute for Cancer Research, Toronto, Canada
James Lopez, Ontario Institute for Cancer Research, Toronto, Canada
Jon Eubank, Ontario Institute for Cancer Research, Toronto, Canada
Justin Richardsson, Ontario Institute for Cancer Research, Toronto, Canada
Leonardo Rivera, Ontario Institute for Cancer Research, Toronto, Canada
The Pan-Canadian Genome Library, Canadian Centre for Computational Genomics, Montreal, Canada
Melanie Courtot, Ontario Institute for Cancer Research; Dept. of Medical Biophysics, Computer Science Dept., University of Toronto, Canada

**Presentation Overview:**
Integrating heterogeneous data across institutions remains a barrier to collaborative research. Traditionally, data harmonization requires extensive custom programming that is difficult to scale, maintain, and audit. We present a configuration-driven framework that decouples transformation logic from implementation. Domain experts can declare complex data transformation rules in human-readable configuration files without deep programming knowledge. Key features include: (1) a five-stage pipeline for common operations: preprocessing, transformation, post-processing, validation and summary, (2) declarative configuration for concrete mapping logics, and (3) preservation of source labels for bidirectional traceability. It employs modular design that can be adapted to diverse source and target schemas beyond PCGL. Validated on the HostSeq dataset, the framework harmonized 500+ REDCap variables into 15 standardized PCGL entities including transforming 360+ multi-select variables from wide format into normalized records with standardized ontology codes assigned. The configuration-driven approach significantly reduces development time while built-in quality assurance improves data harmonization quality. This framework ensures the reproducibility and auditability required for regulatory-grade research. To address ontology mapping bottlenecks, we are prototyping a LLM-based AI layer to recommend ontology codes and generate configurations directly from data dictionaries. This architecture accelerates research readiness by lowering technical barriers to high-quality data harmonization across research ecosystems.

## B-089: Systematic Mapping of RNA Modifications Across Tissues andSpecies

**Authors List:**
**Alex Sizykh**, University of Manitoba, Canada
Jemaneh Habtewold, University of Manitoba, Canada
Elly Wu, University of Manitoba, Canada
Athanasios Zovoilis, University of Manitoba, Canada

**Presentation Overview:**
Post-transcriptional modifications of RNA have emerged as keyl modulators of transcriptome function, influencing gene expression, RNA stability, subcellular trafficking, and translational efficiency. Here, we leverage Oxford Nanopore direct RNA sequencing to profile four principal epitranscriptomic marks-N6-methyladenosine (mA), N4-acetylcytidine (ac4C), pseudouridine (4), and 5-methylcytidine (mC) - within human and murine tissues at single-nucleotide resolution. We systematically map the distribution of these modifications across diverse RNA biotypes and organs, with the overarching aim of creating a comprehensive reference atlas for both species. This poster presents the inaugural instalment of the project: an in-depth interrogation of the most abundant and extensively characterised mark, m6A. Our data recapitulate canonical mA enrichment patterns reported previously and further demonstrate that deposition of this modification is highly context-dependent, underscoring the need for continued mechanistic investigation. We also present the results of our analysis of interplay between m6A and alternative RNA splicing.

## B-091: Lessons learned in optimizing plant breeding workflows for better breeding outcomes

**Authors List:**
**Carolyn T. Caron**, University of Saskatchewan, Canada
Lacey-Anne Sanderson, University of Saskatchewan, Canada
Ruobin Liu, University of Saskatchewan, Canada
Reynold Tan, University of Saskatchewan, Canada
Vidusha Wilpita, University of Saskatchewan, Canada
Laura Jardine, University of Saskatchewan, Canada
Kirstin E. Bett, University of Saskatchewan, Canada

**Presentation Overview:**
With seventeen years of experience in developing and managing a breeder?focused web portal for pulse crop researchers, our team has gained valuable insight into the data?management challenges faced by small and underfunded plant breeding programs. Our current platform integrates germplasm passport, phenotypic, and genotypic data to support key breeding activities, which include parent selection, crossing?block design, and performance?trial analyses. A dedicated curator has been invaluable, by 1) ensuring that pre?breeding germplasm?characterization data remain current, 2) assisting our researchers in removing specialist-bias from experiment documentation, and 3) linking disparate datatypes through common identifiers. Tools such as BLAST, JBrowse and Helium are integrated within the portal to streamline pre-breeding research, and linking these data to breeding material strengthens continuity between research and breeding. Together, these approaches?contribute to the long-term success of agricultural?plant?breeding, enhancing crop resilience, productivity, and?nutritional value in the face of?climate?and environmental challenges. In our future work, we aim to develop databases that support additional plant breeding programs and integrating more breeding-specific tools such as genomic selection to further strengthen the link between pre-breeding and breeding. We welcome the opportunity to share our development process and data-management workflow, and our commitment to FAIR data principles.

## B-093: A New Star in the Galaxy: Canada is Joining the UseGalaxy.* International Federation

**Authors List:**
**Pierre-Étienne Jacques**, Université de Sherbrooke, Canada
Charles Coulombe, Université de Sherbrooke, Canada
Carol Gauthier, Université de Sherbrooke, Canada
John Davis, Johns Hopkins University, United States
Tannistha Nandi, University of British Columbia, Canada
Ata Roudgar, Simon Fraser University, Canada
Jiarui Li, University of Alberta, Canada

**Presentation Overview:**
UseGalaxy.ca is Canada's public instance of the international Galaxy Project, a web-based platform democratizing access to advanced research computing through discipline-specific interfaces and workflows. Galaxy enables researchers without computational expertise to perform complex analyses requiring high-performance computing (therefore perfect for training life scientists!). The platform supports FAIR principles through accessible, reproducible, and transparent research practices. Since its 2024 launch on Alliance cloud infrastructure, UseGalaxy.ca has rapidly grown to serve >1,000 active users (predominantly Canadian researchers) who have executed ~150,000 analytical jobs utilizing more than 42 TB of storage. The platform provides access to thousands of tools and workflows spanning bioinformatics, image analysis, machine learning, chemistry, astronomy, climatology, social sciences, and beyond, supported by a vibrant international community of developers and trainers. UseGalaxy.ca joins a global federation including major instances in the United States, Australia, and Europe, collectively serving over 600,000 registered users. The Canadian instance is managed by a dedicated technical team integrated with the Alliance and supported by international Galaxy developers. This national service demonstrates how strategic infrastructure investments enable cutting-edge, accessible research capabilities for Canadian scientists across all disciplines.

## B-095: Integrated Breast Tumor and Placenta Organ-on-a-Chip Modelto Define the Molecular Profile of Pregnancy-AssociatedBreast Cancer

**Authors List:**
**Yasmin Abdelkader**, University of Toronto, Canada

**Presentation Overview:**
Pregnancy associated breast cancer (PABC) is the most common cancer in pregnancy in which 4% of all breast tumors are coincident with pregnancy with a steady-state increase in occurrence in Canada and globally. The lack of complete molecular profiling of PABC condition raises many challenges for developing safe treatment. Therefore, our objective is to develop a biomimetic microenvironment to investigate the molecular profile of the placenta and tumor in PABC using organ-on-a-chip technology coupled with multi-omics and bioinformatic approaches. We developed an integrated dynamic breast cancer-on-a-chip and placenta-on-a-chip model using MCF7 and BeWo cells, respectively where cells were collected for RNA sequencing and media were analyzed for secreted proteins. Our data showed that the dynamic conditions resulted in decreasing barrier function, adhesion, and tight junction formation in BeWo cell line in integrated dynamic conditions compared to static controls. The breast tumor also showed increased ECM formation, characterized by upregulation of collagen complex (COL4A1, COL4A2) and epithelial-mesenchymal-transition (EMT-related genes) along with activation of PI3K/Akt signaling pathway. Our study introduces an innovative approach for characterizing the molecular profiles of PABC-associated cell lines through the integration of an organ-on-a-chip model and multi-omics analysis, providing a more precise representation of the PABC pathological state.

## B-097: Resolving HIV-1 Transmission Risk Structure in Genetic Clusters with Community Detection

**Authors List:**
**Paula Magbor**, Western University, Department of Pathology and Laboratory Medicine, Canada
Art Poon, Western University, Department of Pathology and Laboratory Medicine, Canada

**Presentation Overview:**
Virus transmission rates are often studied by generating networks linking genetically similar infections. By convention, component clustering (CC) is used to extract subsets of infections from the network. However, CC can obscure the risk heterogeneity within large components, e.g., two cliques connected by a single edge are one component. We investigated the application of community detection (CD) to resolve this issue. We implemented multi-deme compartmental models in TiPS, and simulated HIV-like sequences with Pyvolve for generating networks. The TN93 genetic distance was also applied to 8,593 HIV-1 pol sequences from China, and the resulting pairwise matrix was used to construct networks under different thresholds (0.005-0.045 substitutions/site). We extracted clusters from these networks using seven methods in CDlib, and computed the adjusted mutual information (AMI) between clusters and known risk factors. We obtained higher AMI (0.271-0.297) for CD methods at higher thresholds (TN93=0.035) than CC (AMI=0.165, TN93=0.025) on simulations, demonstrating a benefit to partitioning large components. For HIV-1 data, AMI was also higher for CD (0.137-0.145) than CC (0.042) at higher TN93 thresholds. CD also showed higher concordance with sampling locations by province, consistent with latent risk structure.

## B-099: EcoPopDL-GP: Environmental-Aware and Population-InformedGenomic Prediction using Deep Learning and ChromoMap

**Authors List:**
**Thulani Hewavithana**, University of Saskatchewan, Canada
Isobel Parkin, Agriculture and agri-food canada, Canada
Lingling Jin, University of Saskatchewan, Canada

**Presentation Overview:**
Predicting traits shaped by genotype×environment (G×E) interactions remains challenging in genomic prediction, especially for highly polygenic traits. Although deep learning (DL) captures non-linear effects, many approaches ignore chromosome organization, haplotype-block structure, SNP positional relationships, and population structure, limiting generalization. We present EcoPopDL-GP, a multimodal DL framework integrating genomic structure and environmental dynamics. Ordered SNPs are encoded in a chromosome-aware 3D tensor (“ChromoMap”) preserving positional relationships and haplotype-block context, with optional gene- and transposable element-derived channels providing functional genomic context. A position-aware Transformer aggregates genome-wide patterns. Temporal G×E effects are modeled using an LSTM on windowed weather data, while location, year, and population are encoded as learnable embeddings capturing systematic field-trial differences. Genomic and environmental embeddings are fused via an interaction module, and a gated dual-branch head balances additive and interaction effects to model polygenic and simpler traits. Ablation across five model variants shows chromosome-aware encoding, gene/TE embeddings, environmental modeling and the interaction module each improve performance, yielding the strongest gains for polygenic traits. Benchmarking against GBLUP, Bayesian ridge regression, random forest, XGBoost, learnMET, and DeepGxE across chickpea and rice panels (four traits), EcoPopDL-GP achieved $R^2$=0.50-0.62, representing 11-32% relative improvement over the strongest baselines, demonstrating improved generalization under complex G×E scenarios.

## B-101: Multiomics approaches to characterize host-microbeinteractions in pediatric IBD patients with comorbid anxiety

**Authors List:**
**Daniel Fry**, University of Toronto, Canada
John Parkinson, Hospital for Sick Children, Canada

**Presentation Overview:**
Inflammatory bowel diseases (IBDs) are a complex, multifactorial set of diseases, including Crohn's Disease and Ulcerative colitis. IBD incidence is increasing globally, especially among children: very-early (6 years of age) onset IBD, is the fastest growing patient subpopulation. Research has indicated that the bi-directional influence of the brain and gut microbiome is important in IBDs – a notable example is anxiety, whose co-occurrence with IBDs increases disease severity and treatment resistance. The exact host-microbe interactions in IBD patients with anxiety remain poorly understood, especially in pediatric patients. To address this, we have recruited a cohort of pediatric patients co-diagnosed with IBDs and anxiety and performed deep read metagenomics, meta-transcriptomics, and metabolomics on stool samples. We will interrogate these data to assess differences between our anxiety and control groups to develop a comprehensive picture of how anxiety interacts with the IBD gut. Finally, we will use flux balance analysis to build comprehensive metabolic models for each patient and perform in silico screening to identify nutritional supplements that may modulate the effects of anxiety on the IBD gut microbiome. This research will help advance our understanding of the interactions between anxiety and IBDs and help uncover possible avenues for precision IBD treatments.

## B-103: Dissecting the regulatory mechanism governing SINE Alu RNAexpression in response to cellular stress

**Authors List:**
**Alyona Gerasimova**, University of Manitoba, Canada
Parisa Rayomand, University of Manitoba, Canada
Riya Roy, University of Manitoba, Canada
Travis Haight, University of Manitoba, Canada
Athanasios Zovoilis, University of Manitoba, Canada

**Presentation Overview:**
Alu elements are primate-specific short interspersed nuclear elements (SINEs) comprising >10% of the human genome and are transcribed into ~300-nt non-coding RNAs. During stress, sense Alu RNAs accumulate, repress transcription by binding RNA polymerase II, and undergo fragmentation, but their antisense transcripts are poorly characterized. This project aims to characterize the interplay between sense and antisense Alu RNAs during cellular stress. To address this, we used heat shock as a cellular stress model. HeLa cells were exposed to heat shock and strand-specific qPCR, and stranded RNA-seq of short (?200 nt) and long RNA fractions was applied. To assess directional coupling, we also selectively targeted sense or antisense Alu RNAs using locked nucleic acid (LNA) oligonucleotides. Our findings revealed a stress-responsive Alu regulatory circuit and identified Alu transcripts as candidate modulators of transcriptional homeostasis, with potential relevance to disease states involving disrupted stress signaling and non-coding RNA regulation.

## B-105: iMicroSeq: Integrated resources supporting clinical andenvironmental sequence data

**Authors List:**
Robert Delatolla, University Of Ottawa, Canada
Fiona Sl Brinkman, Simon Fraser University, Canada
Emma Griffiths, Simon Fraser University, Canada
Mélanie Courtot, Ontario Institute for Cancer Research, Canada
Erin E Gill, Simon Fraser University, Canada
Math'Ieta Alatini, Council of Yukon First Nations; One Yukon Coalition, Canada
Guillaume Bourque, McGill University, Canada
José Héctor Gálvez López, Canadian Centre for Computational Genomics, Canada
Wlliam Wl Hsiao, Simon Fraser University, Canada
Michael Parkins, University of Calgary, Canada
Casey Rj Hubert, University of Calgary, Canada
**Baofeng Jia**, Simon Fraser University, Canada
Imicroseq Consortium Leadership Table, iMicroSeq Consortium, Canada
Ontario Wastewater Surveillance Consortium, Ontario Wastewater Surveillance Consortium, Canada
Mohamed Kharbouch, Ontario Institute for Cancer Research, Canada
Carolina Koutras, One Yukon Coalition, Canada
Johanie Gilbery, Simon Fraser University, Canada
Leonardo Rivera,

**Presentation Overview:**
iMicroSeq.ca (inclusive, integrative Microbial Sequencing) is a national initiative grounded in the principle of complementing clinical microbial sequence data with environmental DNA/RNA water-based analysis to wholistically Learn from Water to improve the health of people, animals, and the environment. Here we describe interlinked resources, including the iMicroSeq Data Portal for sequence and qPCR data, based on the Overture Framework, and tools to support modelling and visualization of the data. Our work is rooted in a data governance framework that supports FAIR Principles, data security, privacy, recognition of submitters, robust data standards/metadata specifications, government-academic collaboration, and data governance transparency. The associated ChùNet.ca project is supporting development of an Indigenous-led local knowledge and data sharing network for monitoring that reflects community preferences. An example of more integrative analysis made possible through using iMicroSeq will also be presented. Collectively, we describe an inclusive framework that aims to enable robust microbial/eDNA sequence data linkage, discovery, and analysis, with relevant environmental/other data to advance equitable water-based monitoring of pathogens and other microbes within a One Health framework. Results of this work supports identifying predictors of disease trends, improving forecasting of community-level risk, and informing public health policy and interventions.

**B-107: MicroNucML: A machine learning approach for micronucleisegmentation and the refinement of nuclei-micronucleirelationships**

**Authors List:**
**Nadejda B. Boev**, University of Toronto; University Health Network, Canada
Yukai Wang, University of Toronto; University Health Network, Canada
Ulises O. Garcia, University of Toronto; University Health Network, Canada
Kate M. MacDonald, University of Toronto; University Health Network, Canada
Shane M. Harding, University of Toronto; University Health Network, Canada
Sushant Kumar, University of Toronto; University Health Network, Canada

**Presentation Overview:**
Micronuclei (MN) are structures containing small fragments of DNA, arising from mitotic errors or failed DNA repair attempts. Therefore, MN serve as markers of genomic instability and are typically quantified either manually or through threshold-based methods, which can be tedious and inaccurate, leading to varying degrees of success and throughput. By employing a two-phase labeling approach that utilizes polygon and brush segmentation, along with refinement using SAM2, we developed a high-quality MN segmentation tool. Subsequent data augmentation, which captured heterogeneity in image quality and color diversity, enabled us to train a generalizable Mask-RCNN model optimized for small object detection, achieving state-of-the-art performance in MN detection. Finally, we applied our model to immunofluorescence data obtained from cell lines exposed to DNA damage conditions to gain biological insights into MN dynamics and their role in inducing genome instability. In summary, this work establishes an accessible resource for systematically studying genome instability with significantly greater fidelity and sensitivity, enabling previously unresolved insights into damage biology.

## B-109: Early Genomic Insights into Mycobacterium avium fromParticipants and Their Household Water

**Authors List:**
**Idowu B. Olawoye**, Department of Microbiology and Immunology, Western University, London, ON, Canada
Teresa Kumblathan, Public Health Ontario, Toronto, ON, Canada
David Alexander, Cadham Provincial Laboratory, Winnipeg, MB, Canada
Melissa Richard-Greenblatt, Hospital for Sick Children, Toronto, ON, Canada
Angela Ma, Public Health Ontario, Toronto, ON, Canada
Sarah K. Brode, University Health Network, Toronto, ON, Canada
Theodore K. Marras, University Health Network, Toronto, ON, Canada
Jennifer L. Guthrie, Department of Microbiology and Immunology, Western University, London, ON, Canada

**Presentation Overview:**
Mycobacterium avium (Mav) is the leading cause of nontuberculous mycobacteria pulmonary disease in Canada, with increasing prevalence worldwide. Polyclonal and ongoing environmental Mav infections likely contribute to treatment challenges. To investigate these issues, we initiated a prospective study and present preliminary results from the first ten participants. Whole genome sequences were obtained for 25 clinical and 9 household water isolates. Analyses included core-genome multi-locus sequence typing (cgMLST) using chewBBACA, core-genome short nucleotide variant (SNV)-based phylogenies, pan-genome analysis, and plasmid network analysis with pling and MOB-suite to assess genomic relatedness and plasmid sharing. Preliminary cgMLST and SNV analyses revealed four major clusters, largely geographically structured. Within clusters, clinical isolates from two individuals were closely related (? 7 SNVs) to their household water isolates, suggesting potential source attribution. Three patients exhibited evidence of polyclonal infection, with clinical isolates differing by both cgMLST (50 allelic distance) and core-genome SNV (> 17 SNVs). Plasmid network analysis revealed three containment networks indicating interconnectedness between environmental and clinical isolates. Shared accessory gene signals between clinical and environmental isolates further support potential environmental–clinical linkages. Together, these preliminary findings show that integrated genomic analyses enable high-resolution discrimination of Mav lineages, environmental–clinical relatedness, and polyclonal infection.

**B-111: Towards a Platform for Schema-Level Harmonization: DataCatalogue, Schema Generation, and Automated Mapping**

**Authors List:**
**Ivan Gill**, CIDGOH, Canada
Jun Duan, CIDGOH, Canada
Damion Dooley, CIDGOH, Canada
William Hsiao, CIDGOH, Canada

**Presentation Overview:**

Datasets developed without using community or consensus standards often lack consistency, interoperability, and reusability. Structured data schemas address these issues by explicitly defining data requirements, but are frequently developed in isolation. Existing data portals typically prioritize data retrieval and offer limited support for schema-level operations such as discovery, comparison, and reuse. To address this gap, we developed a data catalogue that indexes and queries schema-level information and includes integrated schema similarity search to identify related or overlapping schemas. To enable schema-level harmonization during data entry and curation, we enhanced our previously published, browser-based spreadsheet editor DataHarmonizer. A built-in LinkML editor allows users to generate and modify schemas directly in the interface, facilitating reuse and extension of existing schemas. We further extended validation by integrating LLM-based automated mapping of user-entered values to controlled vocabularies defined in schemas, reducing manual effort and accelerating harmonization. We evaluated the LLM-based mapping approach using diverse input variants, including spelling differences, abbreviations, and conceptually related phrases. While LLMs generally mapped inputs accurately, they exhibited high latency and reduced reproducibility when multiple valid mappings were possible. These findings underscore key considerations for building a scalable, semantically informed within-schema harmonization pipeline that supports consistent data integration across projects.

## B-113: PlastBurstAlign: Scalable extraction, alignment, andannotation validation across thousands of plastid genomes

**Authors List:**
**Thanina Hamitouche**, Fort Hays State University, United States
Michael Gruenstaeudl, Fort Hays State University, United States

**Presentation Overview:**
Plastid phylogenomics has been a key approach in plant evolutionary studies for a decade, yet extracting homologous regions across large sets of plastid genomes has remained technically challenging. In particular, efficient workflows that address annotation inconsistencies, structural genome complexity, and the automated compilation of sequence alignments are needed for datasets comprising thousands of plastid genomes. Here, we present PlastBurstAlign, a Python-based software package designed to automatically extract homologous regions and perform multiple sequence alignment across extensive plastid genome collections. The tool implements an automated pipeline that identifies and groups homologous genes, introns, and intergenic spacers while detecting annotation discrepancies before alignment. It specifically addresses major challenges in plastome analysis, including mosaic gene organization, annotation errors requiring automated filtering, the need for distinct alignment strategies for coding versus noncoding regions, scalability through task-level parallelization, and flexible exclusion of user-specified regions. Benchmark analyses demonstrate that PlastBurstAlign enables scalable extraction, annotation validation, and alignment of hundreds to thousands of plastid genomes within practical computational timeframes compared to existing approaches. By improving accuracy, automation, and scalability in plastid genome alignment, PlastBurstAlign facilitates robust plant phylogenomic analyses, supporting large-scale agricultural and evolutionary research.

## B-115: Identification of Epigenetic Subtypes in Follicular Lymphoma

**Authors List:**
**Rajesh Detroja**, University Health Network (UHN), Canada
Robert Kridel, University Health Network (UHN), Canada

**Presentation Overview:**
Follicular lymphoma (FL) is a clinically heterogeneous B-cell malignancy, with a subset of patients experiencing early progression or histologic transformation into diffuse large B-cell lymphoma (DLBCL). To better understand the epigenetic underpinnings of this clinical variability, we performed genome-wide DNA methylation profiling on 338 primary FL samples. Unsupervised clustering revealed two distinct epigenetic subtypes—iFL (indolent-like FL) and aFL (aggressive-like FL). Biologically, aFL was characterized by greater global hypomethylation and focal hypermethylation in CpG-dense regions, increased proliferative history, and a higher burden of copy number alterations and mutations, molecularly aligning it with aggressive DLBCL. Clinically, aFL was associated with progression of disease within 24 months (POD24) and increased risk of histologic transformation. Furthermore, progression-free and overall survival were inferior. In multivariable analysis, DNA methylation alterations emerged as an independent predictor of transformation risk. Our findings demonstrate that DNA methylation patterns capture both biological heterogeneity and clinical risk in FL, establishing a foundation for incorporating methylation-based classification into risk stratification and personalized therapeutic strategies.

## B-117: Interpretable Prediction of Polypharmacy Side Effects via Hypergraph Neural Networks

**Authors List:**
**Sadra Hakim**, School of Computer Science, University of Windsor, Canada
Luis Rueda, School of Computer Science, University of Windsor, Canada
Alioune Ngom, School of Computer Science, University of Windsor, Canada

**Presentation Overview:**
Polypharmacy, the concurrent use of multiple drugs, frequently leads to adverse drug events. Despite the emergence of various predictive models, most remain black boxes that lack the biological insight necessary for clinical utility. We present an interpretable multi-modal framework for polypharmacy side-effect prediction that represents each drug–drug–side-effect event as a 3-uniform hyperedge, consistent with the non-causal, co-occurrence nature of pharmacovigilance-derived benchmarks. Our approach integrates chemical, semantic, and biological information by encoding drugs from SMILES using a pre-trained molecular language model, side effects using biomedical concept embeddings, and proteins using GO-Slim functional annotation vectors. Additionally, biological context is incorporated through a graph of drug–target and protein–protein interactions to learn graph-informed drug and protein representations, which initialize a hypergraph neural network that propagates information across observed triples. A lightweight decoder then scores candidate associations. On the Decagon benchmark, the model achieves competitive performance (0.874 AUROC) and remains robust under a cold drug-pair split (0.837 AUROC). Further, to interrogate model decision-making, we apply Integrated Gradients to quantify protein-level contributions for individual predictions and conduct targeted case studies that recover biologically plausible signals. Overall, this work couples accurate prediction with mechanistically informative attributions, providing a transparent tool for hypothesis generation.

## B-119: Improving Trustworthiness in Deep Unsupervised Taxonomic Assignment

**Authors List:**
**Dylan Lewis**, University of Guelph, Canada
Madelin Stueck, University of Guelph, Canada
Gurjit Randhawa, University of Guelph, Canada
Lila Kari, University of Waterloo, Canada
Kathleen Hill, Western University, Canada

**Presentation Overview:**
In the field of taxonomic assignment, the lack and unreliability of labels can necessitate unsupervised methods. Relying solely on intrinsic data patterns, unsupervised approaches often underperform relative to supervised approaches. DeLUCS, and its successor iDeLUCS, were developed to address this, using ensembles of unsupervised deep learners to capture more complex patterns than simple learners. These methods, however, augment their datasets with synthetic data, calling into question whether learned patterns are real, or artificially derived; since the ensembled deep learners are black boxes, further obscured by their ensembling process, this remains unanswered. This study replaces synthetic-mutation augmentation with sequence-fragment augmentation, and evaluates performance across eleven datasets comprising over 14,000 sequences and more than one billion base pairs. Fragment-based augmentation achieves comparable performance to DeLUCS and iDeLUCS, and outperforms baseline clustering algorithms by up to 46%. The performance differences between our methods and baseline approaches were statistically significant (p 0.05). We further analyze the DeLUCS and iDeLUCS methods, revealing substantial inaccuracies in various facets (including reported clustering performance), which are fixed in our methods. Our findings underscore the need for rigorous and transparent evaluation practices, and provide more trustworthy and reproducible methods for unsupervised taxonomic assignment.

**B-121: A Machine Learning Framework for Accurate Classification ofSNP Genotype Data in Mus musculus and Ovis dalli**

**Authors List:**
**Spandana Chereddy**, School of Computer Science, University of Guelph, Canada
Dylan Lewis, School of Computer Science, University of Guelph, Canada
Gurjit S. Randhawa, School of Computer Science, University of Guelph, Canada
Kathleen A. Hill, Department of Biology, Western University, Canada
Lila Kari, School of Computer Science at the University of Waterloo, Canada

**Presentation Overview:**
Classifying an organism's genetic identity using machine learning (ML) enables rapid population identification amid accelerating genomic data generation. Genotype data, capturing major homozygous (AA), minor homozygous (BB), or heterozygous (AB) alleles across thousands of polymorphic loci, provides a scalable framework for population-level classification. Machine Learning with Digital Signal Processing (MLDSP) was applied to array-based and sequencing-based single-nucleotide polymorphism (SNP) genotype datasets. Supervised classification was conducted on 840 Mus musculus samples genotyped at 493,290 loci per individual and categorized by breeding strain (inbred, outbred, hybrid, recombinant). Additionally, 299 thinhorn sheep (Ovis dalli) sequenced at 9,536 loci per individual were classified by geographic region. Supervised MLDSP achieved up to 98.1% accuracy in mice and 85.1% accuracy in sheep, with sheep classes reflecting geographic clustering in Euclidean space. Unsupervised ML approaches were also applied. Without prior labels, clustering revealed strong intrinsic structure, achieving over 90% alignment with known strain and geographic groupings. These findings demonstrate that SNP genotype patterns alone contain sufficient signal for robust population differentiation. Together, supervised and unsupervised ML combined with digital signal processing provide rapid, technology-agnostic tools for genotype classification and future genotype–phenotype modeling.

## B-123: Benchmarking and optimization of NanoVar with transposable element annotation for long-read structural variant analysis

**Authors List:**
**Asmaa Samy**, Division of BioMedical Sciences, Faculty of Medicine,Memorial University of Newfoundland, St. John's, NL, Canada, Canada
Touati Benoukraf, Division of BioMedical Sciences, Faculty of Medicine,Memorial University of Newfoundland, St. John's, NL, Canada, Canada

**Presentation Overview:**
Structural variants (SVs) are large genome alterations that are diverse in size and type, including unbalanced, balanced, and complex variants. SVs are a major source of genetic diversity and disease predisposition. Accurate characterization of SVs, including transposable elements (TEs), remains challenging, but long-read sequencing technologies have improved detection despite high sequencing error rates and low throughput. With the rapid advancement of sequencing technologies, many SV callers have been developed, including NanoVar, an optimized SV caller for low-depth long-read sequencing data. We performed benchmarking and optimization of NanoVar across sequencing platforms and reference genomes for robust SV analysis. The latest release integrates NanoINSight, a feature that annotates non-reference repeat element insertions detected by NanoVar. Benchmarking with the Genome in a Bottle dataset guided algorithm parameter tuning and improved detection performance. The results also showed that the choice of reference genome significantly impacts read mapping and SV calling. These results demonstrate the importance of optimization and evaluation of long-read SV analysis workflows, contributing to the advancement of SV detection in genomic studies.

## B-125: The Centre Québécois de Données Génomiques (CQDG): a platform for data harmonisation, sharing, and interpretation.

**Authors List:**
**Daniel Tremblay-Sher**, Centre de recherche Azrieli, CHU Sainte-Justine, Université de Montréal, Canada
Georgette Femerling, Centre de recherche Azrieli, CHU Sainte-Justine, Université de Montréal, Canada
David Morais, Centre de recherche Azrieli, CHU Sainte-Justine, Université de Montréal, Canada
Jeremy Costanza, Centre de recherche Azrieli, CHU Sainte-Justine, Université de Montréal, Canada
Vincent Ferretti, Centre de recherche Azrieli, CHU Sainte-Justine, Université de Montréal, Canada

**Presentation Overview:**
Data sharing platforms are critical for multi-omic studies not only to promote FAIR principles and meet funding agency requirements, but also to provide access to compute resources as well as to large, searchable datasets. The CQDG provides a secure infrastructure for storing, sharing, and interpreting research data. The platform is hosted as part of the Secure Data for Health (SD4H) project, with Montréal-based physical infrastructure managed by Calcul Québec. Data is submitted via a user-friendly web portal and harmonized using an ontology-based data dictionary. Specialized files are deposited in an S3 compatible staging environment, after which a rigorous quality control pipeline evaluates alignment metrics in accordance with GA4GH standards, as well as variant-level metrics, pedigree consistency and relatedness, and gene-level coverage. The platform extracts, annotates, and prioritizes genomic variants using a Nextflow pipeline integrating VEP and Exomiser, and publishes the results in an anonymized, searchable database. A growing suite of tools for variant interpretation based on pathogenicity and frequency as well as other analyses and visualizations is being integrated. CQDG users can define virtual cohorts using specific patient and data characteristics. Contributing studies retain granular control over data access (at the participant or file level) at all times.

## B-127: manylatents-omics: A Unified Framework for DimensionalityReduction and Metric-Based Evaluation in Multi-Omics

**Authors List:**
**Matthew Scicluna**, University of Montreal, Canada
César Miguel Valdez Córdova, Quantitative life sciences. McGill university, Canada

**Presentation Overview:**
Dimensionality reduction is fundamental to interpreting high-dimensional biological data, yet standard tools often lack the domain-specific context required to rigorously evaluate latent representations. We present manylatents-omics, a modular extension of manylatents—a comprehensive framework for manifold learning and geometric evaluation. By building on this core infrastructure, manylatents-omics standardizes the inference, analysis, and metric-based evaluation of embeddings across diverse omics modalities. The software implements a unified architecture comprising three core modules: popgen for population genetics, singlecell for transcriptomics, and dogma for central dogma sequences. Methodologically, it includes built-in benchmarking datasets (e.g., PBMC 10X Genomics datasets) and facilitates high-throughput inference using foundation model encoders (e.g., ESM3, Evo2, AlphaGenome) alongside streamlined data loaders for standard formats like PLINK and AnnData. Crucially, the framework prioritizes quantitative evaluation by supplementing the general geometric metrics inherited from the parent package with domain-specific measures, such as admixture preservation. Both packages will be released in early May. By integrating rigorous domain-specific benchmarking with standard encoders, manylatents-omics enables reproducible evaluation and accelerates the application of computational methods to complex biological datasets.

## B-129: Variations in ADAR editing of nonsense mediated decaytargets in PD males and females

**Authors List:**
**Heather Mercer**, University of Mount Union, United States
Helen Piontkivska, Kent State University, United States
Aiswarya Mukundan Nair, Kent State University, United States
Ayesha Tariq, Kent State University, United States

**Presentation Overview:**
Parkinson’s Disease (PD) is a complex disease with multiple phenotypes that vary between individuals, as well as by age and sex. Because only a fraction of PD cases can be tied to genetic variants, it is likely that a complicated interaction between gene expression, hormones, the environment, and modifications to RNA transcripts plays a role in PD pathology and progression. Exploring changes in RNA editing between healthy controls and PD males and females aged 65 years or older enrolled in the Parkinson’s Progression Markers Initiative (PPMI), we analyzed editing through the actions of the adenosine deaminases acting on RNA (ADARs), which may cause nonsynonymous alterations to gene expression products including those that result in nonsense-mediated decay (NMD). Interestingly, the pattern of NMD-related editing affecting transcripts of iron uptake transporter SLC11A2, or DMT1, varied in PD by sex with those specific edits observed in only PD females. Our findings suggest that the dysregulation of ADAR editing may play a role in PD and further, that ADAR editing associated with NMD and genes functioning in NMD-related pathways may be integral to PD pathophysiology, particularly between the sexes.

## B-131: Neurosymbolic artificial intelligence offers a unique approach for discovering biological laws when combined with category theory that is being used to relate different biological variables

**Authors List:**
**Efe Yagiz Aygun**, University of New Brunswick, Canada

**Presentation Overview:**
Bioinformatics is trying to figure out what does the data we have mean. It can get complicated when we do not know whether we are using the right tools. However, it also have the potential to discover new biological laws. Neurosymbolic Artificial Intelligence (AI) combines deep learning with Symbolic AI to resemble human like logical reasoning and have the potential to uncover new biological insights when applied to bioinformatics. Furthermore, when Category Theory is integrated with this approach, it can offer unique methods and tools to bioinformaticians to accelerate research and improve our knowledge. It is because there are many biological data types and variables and we need more in depth study to map their relation to each other. In that regard, Category Theory has the potential to bridge mathematical knowledge and biology, which have been studied theoretically in the past. Therefore, this research is trying to use Category Theory as a starting point to achieve this goal and offer an early exploratory framework insights for applying this idea into biological organization, combining both Neurosymbolic AI and Category Theory, that can be shared with the bioinformatics community.

## B-133: VDJump: a computational tool to detect and characterizetumour-infiltrating lymphocytes in tumour WGS data

**Authors List:**
Haley Pedersen, University of Calgary, Canada
Brooklyn Christopher, University of Calgary, Canada
Grace Mariene, University of Calgary, Canada
**Paul Gordon**, University of Calgary, Canada

**Presentation Overview:**
The tumour microenvironment often contains a diverse mixture of cell populations in addition to malignant cells. This may include tumour-infiltrating lymphocytes (TILs), including B and T cells that have migrated into the tumour and can recognize and kill cancer cells. TILs are a critical component of the immune system's response to cancer. A high concentration of TILs, especially cytotoxic CD8+ T cells, is often associated with better clinical outcomes in many cancers. We have developed VDJump, the first software package that estimates the proportion of lymphocytic cells in tumour samples directly from tumour whole genome shotgun (WGS) data. VDJump estimates are primarily based on reduced depth of coverage in T and B cell receptor genome loci. These loci are subject to “variable-(diversity)-joining”, a.k.a. V(D)J, nuclear DNA recombination events that are specific to lymphocyte development. Ongoing development of VDJump is focused on cell-subtype characterization to further break down the T cell proportion estimate into CD8+ or CD4+ T cells by leveraging public single cell data. Additionally, we are evaluating the performance of VDJump in a large multi-cancer WGS dataset with matched multiplexed immunohistochemistry benchmark data as part of a national data-sharing initiative, including comparisons to existing RNA-based deconvolution methods.

## B-135: RIDA - Result-Informed Data Acquisition: Novel machine learning software to optimize data collection in mass spectrometry-based proteomics

**Authors List:**
Iryna Abramchuk, University of Ottawa, Canada
Yun-En Chung, University of Ottawa, Canada
Aarthie Senathirajah, University of Ottawa, Canada
Maram El-Azouni, University of Ottawa, Canada
Giorgio Freije, University of Ottawa, Canada
Alona Petrova, University of Ottawa, Canada
Jonathan St-Germain, Princess Margaret Cancer Centre, Canada
Jens Decker, Bruker Daltonics GmbH & Co.KG, Germany
Brian Raught, Princess Margaret Cancer Centre, Canada
Jonathan R. Krieger, Bruker Ltd., Canada
Tharan Srikumar, Bruker Ltd., Canada
Theodore Perkins, Ottawa Hospital Research Institute, Canada
**Mathieu Lavallée-Adam**, University of Ottawa, Canada

**Presentation Overview:**
While mass spectrometry is the most sensitive and popular technique to characterize proteins in complex biological samples, it fails to comprehensively identify and quantify all proteins in such samples. Current mass spectrometry data acquisition strategies either collect too much redundant data from abundant proteins and none for the detection of less abundant ones, or acquire very complex data from most proteins in a format making protein identification difficult. These data acquisition strategies share one weakness: they do not use data acquired during the mass spectrometry experiment to adapt and optimize their parameters based on what has been observed so far in the sample. Herein, we present a novel machine learning-based software suite (RIDA: Result-Informed Data Acquisition) that analyzes mass spectrometry data in real-time, as it is acquired, and optimizes data acquisition according to results generated on-the-fly. This suite's tools avoid redundant data collection to improve protein identification (11.9% increase), optimize mass spectrometry parameters to increase signal-to-noise (9.4% more identified peptides), and enhance data collection efficiency, requiring 22.3% and 7.9% less data to identify and quantify proteins, respectively, compared with current methods. Our software-based mass spectrometry data acquisition optimizations improve proteome characterization and our understanding of biological processes in analyzed samples.

## B-137: Multiscale volume electron microscopy of the human liver maps vascular-cellular architecture, organelle dynamics and inter-organelle communication

**Authors List:**
**Cheng Xing**, University of Toronto, Canada
Sonya MacParland, University of Toronto, Canada
Mei Zhen, University of Toronto, Canada
Gary Bader, University of Toronto, Canada

**Presentation Overview:**
The human liver depends on multiscale structural organization from vasculature to cells to organelles to perform its diverse metabolic functions. A unified three-dimensional view linking these hierarchical scales in intact human tissue would be useful for better understanding these levels and how they relate to each other. We present a high-resolution volume electron microscopy reconstruction of human periportal liver tissue acquired by serial block-face scanning electron microscopy. Using a multiscale deep learning approach, we performed automated segmentation across the entire volume, enabling comprehensive annotation of vasculature, cells, and organelles. Quantitative analysis of bile duct architecture revealed coordinated scaling between lumen geometry and cholangiocyte number and size. Sinusoidal capillary branches exhibited distinct structural profiles with differential endothelial coverage. Analysis of 35,790 complete mitochondria identified substantial morphological heterogeneity, with elongated mitochondria displaying preferential endoplasmic reticulum (ER) contacts at narrowing sites, a pattern consistent with ER-mediated fission or fusion. This multiscale reconstruction establishes an ultrastructural reference for the healthy human liver and provides a quantitative framework for investigating hepatic physiology and disease.

## B-139: A transformer-based foundational model for the vaginalmicrobiome

**Authors List:**
**Dollina Dodani**, University of British Columbia, Canada
Natalia Blanco, University of British Columbia, Canada
Mona Aboofazeli, University of British Columbia, Canada
Tushani Pradhan, University of British Columbia, Canada
Aline Talhouk, University of British Columbia, Canada

**Presentation Overview:**
The clinical utility of the vaginal microbiome in gynecology has been limited by sparse and small-scale datasets, which hinder the identification of consistent and generalizable patterns. Foundational models can address this challenge by learning robust, context-aware representations from large, heterogeneous, unlabeled data using self-supervised learning. Here, we present a context-aware, generalized vaginal microbiome foundational model built on a transformer-based architecture and pre-trained on nearly 15,000 diverse samples spanning pregnancy, infections, and gynecologic cancers. Using Community State Types (CSTs), ecologically defined configurations of vaginal microbial communities characterized by dominant taxa, the model captures both compositional and functional context, enabling accurate representation of inter-individual variability. We will demonstrate the model's utility by fine-tuning it for endometrial cancer classification, the second most common gynecologic malignancy globally, using leave-one-study-out validation across seven independent cohorts (N = 424). We hypothesize that CST-based transformer embeddings will outperform conventional approaches, which rely on limited data and are prone to batch effects. By integrating self-supervised learning and domain-specific fine-tuning, this framework will provide a scalable and generalizable approach for vaginal microbiome research, enabling disease-specific predictions even when data are sparse or difficult to collect.

## B-141: Logan: Planetary-Scale Genome Assembly Surveys Life'sDiversity

**Authors List:**
**Artem Babaian**, University of Toronto, Canada
Logan Consortium, ,

**Presentation Overview:**
The breadth of life's diversity is unfathomable, but public nucleic acid sequencing data offers a window into the evolution of genetic diversity across Earth. However the rapid growth and accumulation of sequence data have outpaced efficient analysis capabilities. The largest collection of freely available sequencing data is the Sequence Read Archive (SRA), comprising 39 million datasets or 8×10^16 basepairs. To realize the potential of the SRA, we constructed Logan, a massive sequence assembly transforming short reads into contigs and compressing the data 100-fold, enabling highly efficient petabase-scale analysis. We created Logan-Search, a k-mer index of Logan for free planetary-scale sequence search, returning matches in mere minutes. We use Logan to identify >200 million plastic-degrading enzyme homologs, and validate novel enzymes with catalytic activities exceeding current reference standards. Further, we vastly expand the known diversity of proteins (30-fold over UniRef50), plasmids, P4 satellites, and the recently described Obelisk RNA elements. Logan also enables ecological and biomedical data mining, such as global tracking of antimicrobial resistance genes and the characterization of viral reactivation across millions of samples. Through transforming the SRA, Logan democratizes access to the world's public genetic data and opens frontiers in biotechnology, molecular ecology, and global health.

## B-143: Molecular Dynamics Simulations and Machine Learning ForTherapeutic Strategies Discovery Against MSMP-Driven Cancers

**Authors List:**
**Léopold Quitté**, Département de Médecine Moléculaire, Université Laval, Québec, QC, Canada., Canada
Arnaud Droit, Département de Médecine Moléculaire, Université Laval, Québec, QC, Canada., Canada
Mickael Leclercq, Centre de Recherche du CHU de Québec-Université Laval, Québec, QC, Canada., Canada
Gautier Moroy, Université Paris Cité, CNRS, INSERM, Unité de Biologie Fonctionnelle et Adaptative, F-75013 Paris, France., France

**Presentation Overview:**
The protein MicroSeMinoprotein Prostate associated (MSMP) is overexpressed in several types of disease, notably prostate, ovarian, and breast cancers. This overexpression is critical in tumors resistant to angiogenesis inhibitors therapies, where hypoxia-induced MSMP upregulation drives resistance. MSMP activates a cell receptor (CCR2) and its signaling G-protein, probably promoting tumor growth via a proliferation / differentiation / apoptosis pathway (MAP Kinase). To decipher this mechanism, we employed an advanced computational workflow. The MSMP-CCR2-G-protein complex was modeled using AlphaFold2 and analyzed through molecular dynamics simulations. To evaluate the specific impact of ligands, we compared this assembly to the well-known CCL2-CCR2-G-protein complex and the apo CCR2-G-protein system. These simulations were evaluated using gold-standard computational tools and in-house algorithms, specifically BioDiscML, an AI-based machine learning algorithm. Our results are consistent with experimental data, particularly regarding binding energies and receptor activation patterns. We demonstrated that while MSMP and CCL2 interact with CCR2 through distinct residues, they trigger similar activation mechanisms compared to the apo state. By pinpointing these specific molecular triggers, this study identifies key targets for drug discovery and provides a structural foundation for novel therapeutic strategies to overcome resistance in MSMP-related diseases.

## B-145: NEXA: Agentic Network Exploration for Transparent Querying of Biomedical Knowledge Graphs

**Authors List:**
**Mathieu Arru**, Département de médecine moléculaire, Université Laval & AxeEndo-Nephro, Centre de recherche du CHU de Québec, Canada
Mickaël Leclercq, Département de médecine moléculaire, Université Laval & AxeEndo-Nephro, Centre de recherche du CHU de Québec, Canada
Marie Pier Scott-Boyer, Département de médecine moléculaire, Université Laval & AxeEndo-Nephro, Centre de recherche du CHU de Québec, Canada
Arnaud Droit, Département de médecine moléculaire, Université Laval & AxeEndo-Nephro, Centre de recherche du CHU de Québec, Canada

**Presentation Overview:**
Introduction. Biomedical knowledge graphs (KGs) organize heterogeneous biological data into interconnected networks of genes, diseases, drugs, and pathways, supporting integrative biomedical analysis. However, use of KGs requires expertise in graph query languages, limiting accessibility. Large language models (LLMs) offer natural-language interfaces but often suffer from hallucinations and poor reproducibility. We introduce NEXA, a conversational agent designed for reliable and transparent natural-language querying of biomedical KGs. Methodology. The agent relies on a modular multi-step LLM pipeline that performs entity and relation detection, query generation, execution, and automatic correction. All answers are derived from executable KG queries, with intermediate steps exposed to the user for verification. A web interface enables interaction with multiple KGs, model selection, and exploration of retrieved subgraphs, supporting reproducibility and user oversight. Results. Benchmarking on an Alzheimer's disease knowledge graph shows up to 66% higher accuracy compared to baseline KGQA approaches. Complementary use-case studies illustrate effectiveness in answering complex biomedical queries, alongside gains in precision and interpretability. Conclusion. By tightly coupling LLM reasoning with verifiable KG queries, our approach enables trustworthy natural-language access to biomedical graphs. NEXA is model-agnostic, deployable locally, and suitable for privacy-sensitive biomedical environments, supporting transparent data exploration for research and clinical applications.

## B-147: Scalable Genomics Knowledge Discovery with Direct LLM–Database Integration

**Authors List:**
**David Yuan**, European Bioinformatics Institute, United Kingdom

**Presentation Overview:**
Large Language Models (LLMs) excel at interpreting unstructured text but remain limited in interrogating highly structured data natively. Existing attempts to bridge this gap—such as Model Context Protocols (MCP) and Knowledge Graphs (KG)—introduce substantial architectural complexity and scaling challenges, making them less suitable to the massive volumes of genomics data generated over the past four decades, including over 70 petabytes in the European Nucleotide Archive (ENA). We present a novel approach that integrates LLMs directly with relational databases, eliminating the need for external APIs or pre-materialized graph databases. In this framework, a knowledge graph is generated dynamically using a graph engine. A self-hosted LLM–backed AI agent then interrogates this graph using an AI-friendly Graph Query Language (GQL). As a proof of concept, we modeled antimicrobial resistance (AMR) phenotypes (over 1,500,000 entries) and genotypes (over 100,000 genomes and samples) and successfully queried them using free-text inputs via a web UI. This system enables Graph Retrieval-Augmented Generation (GraphRAG) directly over structured genomics data while remaining fully self-contained, reducing risks of model drift and data leakage. Importantly, it provides a highly intuitive interface for biologists, enabling exploratory queries and discovery beyond the constraints of pre-designed data portals.

## B-149: Machine Learning Reveals an Extreme-Environment k-mer-based Watermark Imprinted Across Eukaryotic Genomes

**Authors List:**
**Monireh Safari**, School of Computer Science, University of Waterloo, Canada
Joseph Butler, Department of Biology, University of Western Ontario, Canada
Gurjit S. Randhawa, School of Computer Science, University of Guelph, Canada
Victoria Kata, University of Western Ontario, Canada
Kathleen A. Hill, Department of Biology, University of Western Ontario, Canada
Lila Kari, School of Computer Science, University of Waterloo, Canada

**Presentation Overview:**
Extreme environments impose selective pressures that drive distinctive adaptations in extremophiles, organisms that thrive under conditions such as extreme temperature, salinity, or pH. Previous studies have shown detectable compositional signatures in prokaryotic genomes linked to extreme environments, enabling classification of bacteria and archaea by environmental niche using only DNA sequence. Whether comparable environmental signatures exist in eukaryotic extremophiles, however, remains unresolved, with important implications for evolutionary biology, biodiversity, and bioengineering. Here, we present a phylogeny-aware computational framework applied to 192 curated eukaryotic whole-genome assemblies spanning three kingdoms, Plantae, Fungi, and Animalia. Genomes are represented using k-mer frequency vectors (k = 3, 6, and 9) derived from randomly sampled genome fragments and analyzed using genus-level cross-validation. We identify a statistically significant, environment-associated component in eukaryotic nuclear DNA, achieving accuracies of 73.23% for extreme temperature, 80.21% for high salinity, and 62.67% for high pH environments (p 0.01). Using structural topic modeling (available as an open-source tool) with genus as a covariate, we identify co-occurring k-mer patterns whose environmental associations persist after controlling for phylogenetic structure. These results are the first evidence of k-mer–based environment-linked signatures in eukaryotic genomes and the first comprehensive, taxonomy-aware analysis distinguishing environment-linked sequence patterns from phylogenetic components.

## B-151: Multi-Omics Integration and Supervised Learning Identifies an Epithelial Signature for Radiotherapy Response in Colorectal Cancer

**Authors List:**
Reuben Kumar, Department of Cancer and Genomic Sciences, College of Medicine and Health, University of Birmingham, UK, United Kingdom
Yujie He, Department of Cancer and Genomic Sciences, College of Medicine and Health, University of Birmingham, UK, United Kingdom
Jiarui Zhou, School of Biosciences and Centre for Environmental Research and Justice (CERJ), University of Birmingham, UK, United Kingdom
S-Cort Consortium, United Kingdom, multiple, United Kingdom
Andrew Beggs, Department of Cancer and Genomic Sciences, College of Medicine and Health, University of Birmingham, UK, United Kingdom
**Deena Gendoo**, Department of Cancer and Genomic Sciences, College of Medicine and Health, University of Birmingham, UK, United Kingdom

**Presentation Overview:**
Colorectal cancer (CRC) is the third most diagnosed cancer globally, and second leading cause of cancer deaths. Radiotherapy is a common treatment, but can demonstrate dangerous side effects and varying patient outcomes. Identifying biomarkers that can effectively predict patient responses to radiotherapy remains necessary for precise treatment of CRC patients. We channelled both unsupervised and supervised approaches to assess radiotherapy response for 233 patients of the S:CORT Consortium. Splitting the cohort, we first integrated matched RNA, CNA, mutation, and methylation profiles of 117 patients using multi-omics factor analysis (MOFA). We identified a radiotherapy signature of 101 biomarkers associated with patients who demonstrate a complete response to radiotherapy, and validated the signature using a random forest classifier on the internal validation dataset, and an independent testing cohort. Our signature effectively predicted treatment outcomes, achieving 89% accuracy with strong discriminatory performance (ROC_AUC = 0.85; PR_AUC = 0.71). Assessing human and murine scRNAseq datasets underscores that the signature is predominantly expressed in CRC epithelial cells, which underpin CRC heterogeneity and cellular diversity. Our identified signature enables pre-treatment identification of CRC patients that are unlikely to achieve a complete response to radiotherapy, thereby sparing these patients from unnecessary radiation exposure and off-target damage effects.

## B-153: Evaluating the influence of annotation and statistical frameworks on pathway analysis results

**Authors List:**
Ritu Verma, University of Guelph, Canada
Jeff Caswell, University of Guelph, Canada
Jeremy Simpson, University of Guelph, Canada
Anthony Mutsaers, University of Guelph, Canada
Sonja Fonfara, University of Guelph, Canada

**Presentation Overview:**
Hypertrophic cardiomyopathy (HCM) is a common acquired cardiac disease in cats and humans, yet its molecular mechanisms is not fully understood. RNA sequencing was performed on left ventricular and atrial tissue from healthy cats and cats with HCM, followed by pathway analysis using multiple tools, including DAVID, g:Profiler, Ingenuity Pathway Analysis (IPA), and Gene Set Enrichment Analysis (GSEA). Considerable variability was observed across pathway analysis approaches. Using default databases, DAVID and g:Profiler identified few mostly generalized biological processes, consistent with limited feline annotation. Using a custom feline background gene database for g:Profiler increased the number and diversity of detected pathways. IPA found numerous canonical pathways; filtering criteria influenced pathway rankings and statistical significance, and potential biologically relevant pathways were not always highly ranked. GSEA, using a custom feline pathway database, identified multiple biologically relevant Gene Ontology terms associated with cardiac structure, metabolism, and remodelling. Across methods, pathway detection and ranking were strongly influenced by database completeness, statistical framework, and filtering stringency, with increased thresholds relevantly reducing pathway detection. These results demonstrate the implication of database content and methodological choices when carrying out pathway analysis, which is of importance for transcriptomic studies in particular when analysing less-annotated genomes.

## B-155: Borderlands Science Automated : Large-scale reverberationthrough crowd augmentation

**Authors List:**
**Parham Ghasemloo Gheidari**, McGill University, Canada
Kai-Hsiang Chang, McGill University, Canada
Bls Development Consortium, McGill University, Canada
Alexander Butyaev, McGill University, Canada
Roman Sarrazin-Gendron, Université du Québec à Montréal, Canada
Jérôme Waldispühl, McGill University, Canada

**Presentation Overview:**
Large-scale studies of the microbiome remain challenging due to several intertwined factors, such as reliance on host parameters and the dependence of methods on upstream analyses, such as multiple sequence alignment and phylogenetic tree construction. In 2020, we launched Borderlands Science, the first citizen science game embedded in an AAA video game, targeting multiple sequence alignment refinement. As a result, we gathered a dataset of more than 135 million human-curated annotations from 4 million players, which we showed improved microbial phylogeny. In parallel, we trained AI agents using behavioural cloning, successfully capturing and reproducing the best human solutions. Now, in our latest project, Borderlands Science Automated, we have automated the citizen science pipeline and integrated AI based on a new paradigm we call “crowd-augmentation,” empowering and upscaling the most important aspect of a citizen science project: the citizens. The final section of our pipeline encapsulates a recently developed beta-diversity method that can uncover biomarkers differentiating microbial communities. Overall, in our preliminary results, we have shown that our pipeline is capable of processing microbial data at a scale previously unattainable and with a quality that enables the discovery of potential biomarkers for several diseases and health statuses.

## B-157: A Convolutional Deep Learning Approach to identify DNASequences for Gene Prediction

**Authors List:**
**Jesus Antonio Motta**, Laval University, Canada

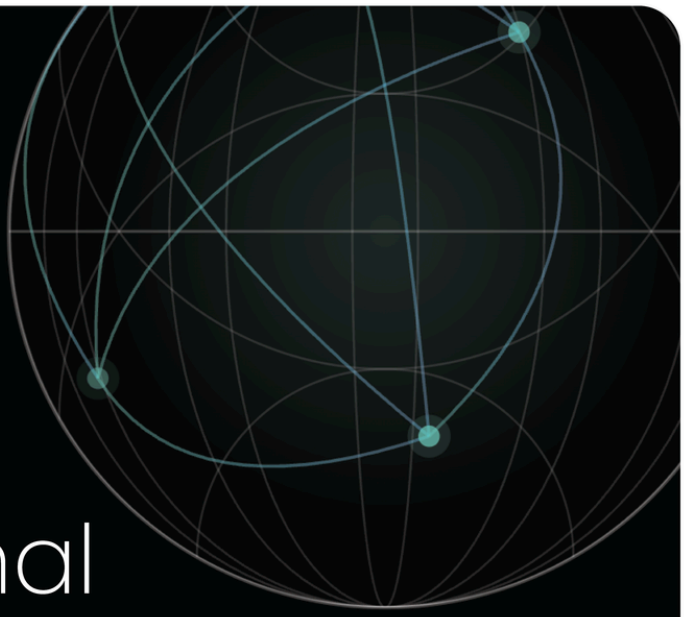
**Presentation Overview:**
We present a high?efficiency machine learning framework for identifying DNA sequences that encode human genes. Using GRCh38 as reference, we compiled training data from multiple curated genomic databases and translated nucleotide sequences into amino acid sequences to construct TF–IDF feature matrices. These representations were used to train a convolutional neural network (CNN) designed to learn discriminative patterns associated with coding regions. Training was performed across all 24 human chromosomes using approximately 36,000 genes and pseudogenes obtained from the HUGO Gene Nomenclature Committee (HGNC). Model performance was evaluated on 24 clinically relevant genes and their surrounding genomic regions. Metrics included precision, recall, F?score, accuracy, and ROC analysis. Our model demonstrates performance that exceeds expectations and achieves state?of?the?art accuracy for gene prediction, as confirmed through direct comparison with AUGUSTUS, our reference baseline. These results highlight the potential of deep learning–based sequence representations for robust gene identification across the human genome.

## B-159: Epigenetic-State–Informed Polymer Modeling RevealsTissue-Specific Principles of 3D Genome Folding

**Authors List:**
**Saman Bazmi**, University Health Network, Canada
Sushant Kumar, University Health Network, University of Toronto, Canada

**Presentation Overview:**
The spatial organization of the genome affects tissue-specific gene regulation by enabling long-range interactions between promoters and enhancers. Beyond static genomic architecture, chromosomal dynamics further influence gene expression through fluctuations in these long-range interactions. To explore the role of 3D genome folding in gene regulation, we integrated experimental Hi-C data, ChromHMM-derived epigenetic state annotations, and coarse-grained molecular dynamics simulations to analyze tissue-specific chromatin structures. We developed an optimized epigenetic-state-informed polymer modeling framework and used it to reconstruct tissue-specific chromatin folding landscapes, characterizing them with contact-probability scaling, compartment order parameters, radius of gyration, and other thermodynamic properties. Our analysis reveals significant tissue-dependent changes in compartment strength, spatial segregation, and local chromatin mixing that cannot be uncovered by ensemble-averaged Hi-C maps. We observed distinct patterns of chromatin compaction and radial redistribution of A/B compartments across genomic regions, reflecting functional specialization and context-dependent regulation. Collectively, these results establish a quantitative, scalable computational framework that links epigenetic state composition to emergent 3D genome structure across tissues, providing a basis for systematic investigation of genome reorganization in development and disease.

CanPath x lifebit



ANNOUNCEMENT

# Canada's National Federated Research Infrastructure.

Lifebit and the Canadian Partnership for Tomorrow's Health (CanPath) have established a national framework that helps researchers collaborate across institutions while data stays secure and under local control — **hosted in Canadian cloud environments.**

1
**Federated**
Data stays with the organization that collected it.

2
**Secure**
Canadian-hosted cloud meeting residency & privacy requirements.

3
**National**
A shared foundation that grows as new partners join.

**350,000+**
Adult participant volunteers across Canada

**1M**
Data points and growing

**178,000+**
Biosamples of DNA source material

CanPath x lifebit

## HOW IT WORKS 01

Each participating cohort retains full governance, oversight, and decision-making authority over its data. Approved researchers can run analyses across multiple sites — without moving sensitive information into a centralized database. An initial research environment is live today, including a synthetic dataset that lets partners safely prepare for collaboration before real datasets are onboarded.

## VOICES 02

> "As more partners join, the benefits grow for researchers, institutions, and ultimately for Canadians."
>
> **Dr. Philip Awadalla**
> National Scientific Co-Director, CanPath

> "Canada does not need another pilot or prototype — we need infrastructure that works at scale, today."
>
> **Maria Dunford**
> CEO, Lifebit

## ABOUT 03

### CanPath

Canada's largest population health study, with **350,000+ volunteer participants.** Hosted by the Ontario Institute for Cancer Research and the University of Toronto's Dalla Lana School of Public Health.

### Lifebit

Global leader in federated, secure, and compliant health data intelligence. Trusted by NIH, Genomics England, and CanPath — managing **275M+ patient records across 30+ countries.**

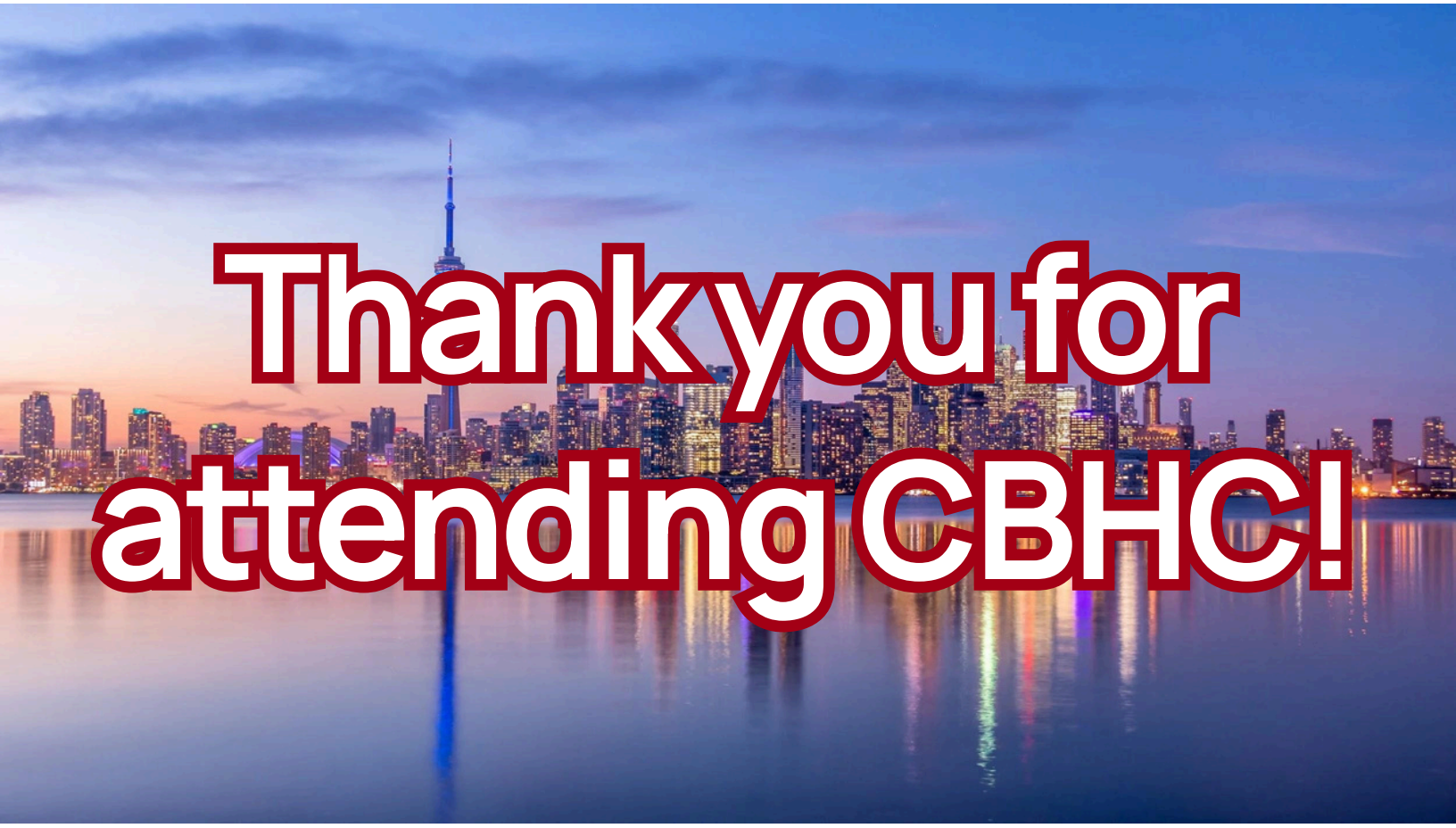
## FOR CANADIAN COHORTS & BIOBANKS 04

**Join the national network — keep your data, share the science.**

Find our teams on-site @CBHC 2026 — let's talk federation.

**Learn more**
canpath.ca
lifebit.ai

Canadian Bioinformatics Hub Conference 2026



# Thank you for attending CBHC!

Canadian Bioinformatics Hub Conference 2026