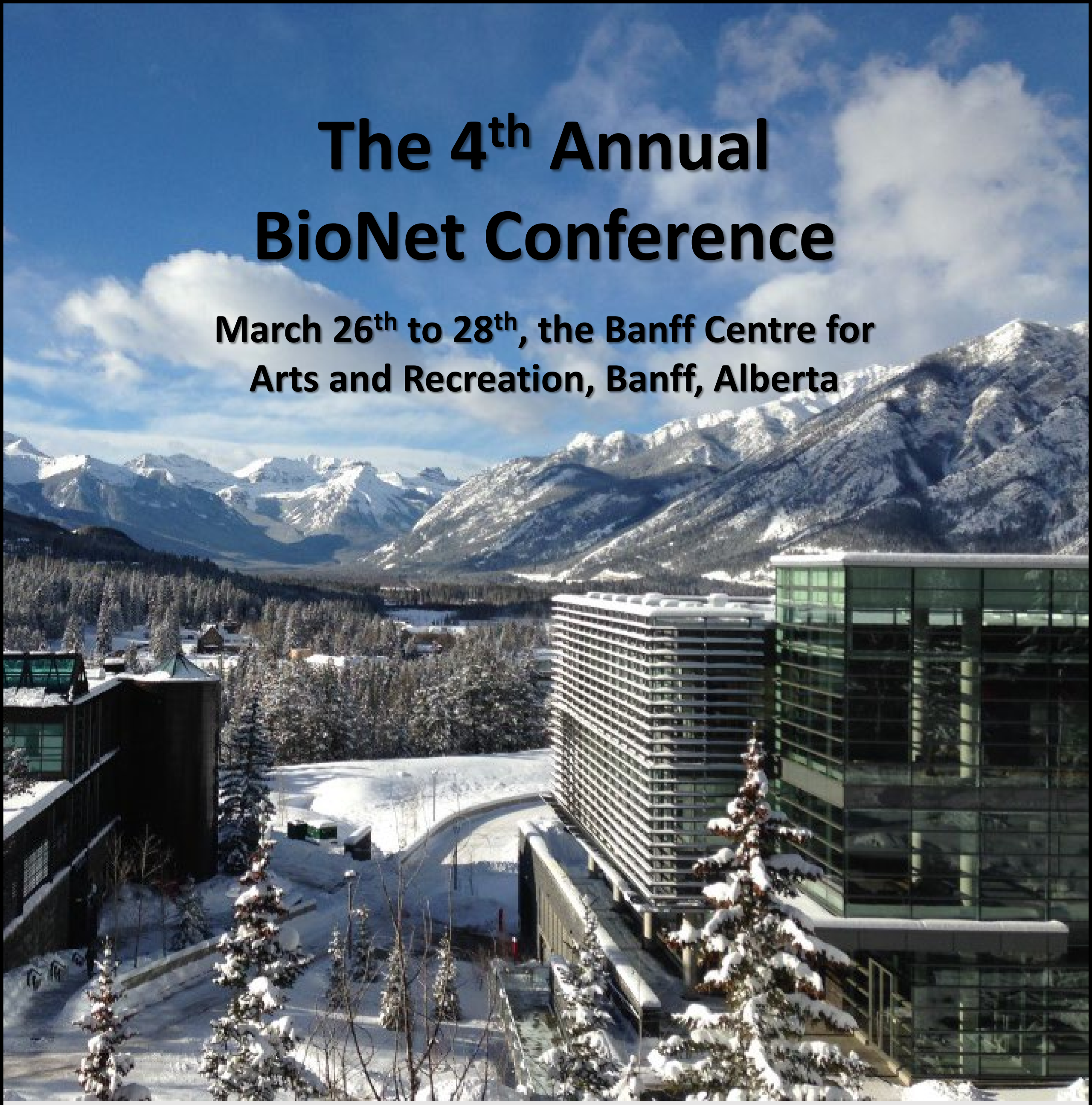
The 4th Annual BioNet Conference

March 26th to 28th, the Banff Centre for Arts and Recreation, Banff, Alberta



Conference Program

Welcome to the 4th Annual BioNet Conference

Dear members of the BioNet community,

On behalf of the organizing committee, we would like to welcome you to the 4th Annual BioNet Alberta Bioinformatics and Omics Conference. BioNet Alberta is a diverse network-based approach to building provincial capacity in Bioinformatics and Computational Biology. We seek to achieve these objectives by fostering an environment of collaboration and cooperation among researchers across a wide variety of disciplines. This conference represents an important physical step in bringing together scientists from institutions across Alberta, the Western provinces and Canada to identify and promote services, platforms and potential partnerships that may be mutually beneficial. We hope you enjoy this year's conference!

4th Annual BioNet Conference Organizing Committee:

Dr. Eric Merzetti Dr. Angeliki Pantazi Dr. Athan Zovoilis

4th Annual BioNet Conference Scientific Organizing Committee:

Dr. Jason de Koning, University of Calgary
Dr. Paul Gordon, University of Calgary
Dr. Emily Herman, University of Alberta
Dr. Elisabeth Richardson, Mount Royal University
Dr. Athan Zovoilis, University of Manitoba

4th Annual BioNet Conference Poster judges:

Dr. Elisabeth Richardson Dr. Ryan Mercer

4th Annual BioNet Conference Student Organizing Committee:

Parisa Rayomand Riya Roy Travis Haight Sani-e-Zehra Zaidi

4th Annual BioNet Conference Sponsors:

PacBio illumina Oxford NANOPORE Technologies

GenomeAlberta Alberta Government GenomeCanada

BIONET BIONET

BioNet Sponsors:

The 4th Annual BioNet Conference
Quick Schedule

Tuesday, March 26th

12:30pm	Registration / Doors Open	KC203 Hallway
1:00pm-1:10pm	Welcome and Opening Remarks	KC203
1:10pm-1:20pm	BioNet Introduction	KC203
1:20pm-1:30pm	Conference Kickoff	KC203
1:35pm-2:45pm	Session 1: Health Omics 1	KC203
2:45pm-3:25pm	Refreshment Break	KC203 Hallway
3:30pm-4:05pm	Opening Plenary: Katrina Ingram	KC203
4:10pm-5:00pm	Session 2: Advanced Comp. Approaches	KC203
5:05pm-5:50pm	Opening Keynote: Dr. Brett Trost	KC203
6:00pm-7:00pm	Supper	KC203
7:00pm-9:00pm	Cocktails and Networking	KC203

Wednesday, March 27th

7:00am-9:00am	Breakfast	Vistas Dining Room
9:00am-9:25am	Plenary: Dr. Steven Jones	KC203
9:30am-10:50am	Session 3: Health Omics 2	KC203
10:50am-11:20am	Refreshment Break	KC203 Hallway
11:20am-12:10pm	Session 4: Microbial Omics	KC203
12:15pm-12:30pm	Sponsor Talk: Illumina	KC203
12:30pm-1:30pm	Lunch	KC203 Hallway
1:30pm-2:15pm	Keynote: Dr. Jérôme Waldispühl	KC203
2:25pm-2:55pm	Session 5: Agricultural Omics	KC203

3:00pm-3:30pm	Refreshment Break	KC203 Hallway
3:30pm-4:45pm	Even Numbered Posters Present	KC205
4:45pm-6:00pm	Odd Numbered Posters Present	KC205

Thursday, March 28th

7:00am-9:00am	Breakfast	Vistas Dining Room
9:00am-9:25am	Plenary: Dr. Michelle Scott	KC203
9:30am-10:35am	Session 6: Health Omics 3	KC203
10:35am-11:05am	Refreshment Break	KC203 Hallway
11:05am-12:10pm	Future Careers Round-Table Discussion	KC203
12:15pm-1:00pm	Awards / Closing Remarks	KC203

4th Annual BioNet Conference – Full Program

Tuesday, March 26th

Check in and badge pickup starting at 12:30pm in Hallway outside Kinnear Centre Room 203

Opening Ceremonies

1:00 – 1:10pm	**Dr. Athan Zovoilis** Academic Lead, BioNet	
1:10 – 1:20pm	**Dr. Eric Merzetti** Network Manager, BioNet Alberta	
1:20 – 1:30pm	**Matt Bryman** Director of Programs, Genome Alberta	

Session 1: Health Omics 1

Chair: Dr. Ryan Mercer

1:35 – 1:50 pm	**Dr. Pinaki Bose, University of Calgary** "Decoding MicroRNA Deregulation: Advancing Prognostication and Therapeutic Strategies in Oral Cancer."	**6**
1:55 – 2:10 pm	**Dr. Britt Drogemoller, University of Manitoba** "The incorporation of automated phenotyping strategies to identify genes and pathways involved in sensory and metabolic hearing loss"	**7**
2:15 – 2:30 pm	**Dr. Arvind Mer, University of Ottawa** "Advancing Precision Oncology: Machine Learning for Drug Response and Biomarker Discovery"	**8**
2:35 – 2:45 pm	**Lilit Antonyan, University of Calgary** "Genetic association analysis of pediatric obsessive-compulsive behaviors, the role of imaging endophenotypes in mediating the relationship between genetic markers and quantitative symptom scores"	**19**
2:45 – 3:25 pm	**Refreshment Break – Kinnear Centre 203 Hallway**	

Opening Plenary

Chair: Dr. Eric Merzetti

3:30 – 4:05 pm	**Katrina Ingram, Ethically Aligned AI** "Data and AI Ethics: A socio-technical perspective"	**3**

Session 2: Advanced Computational Approaches

Chair: Dr. Richard LeDuc

4:10 – 4:25 pm	**Dr. Jason de Koning, University of Calgary** TBD	**9**
4:30 – 4:45 pm	**Dr. Matthew Greenberg, University of Calgary** "A scalable approach to fitting ZINB-WaVE models to single-cell gene expression data"	**10**
4:50 – 5:00 pm	**Kabita Baral, University of Calgary** "Power to detect episodic fitness shift at a small number of sites can be rescued with a covariate-informed branch-site type method"	**20**

Opening Keynote

Chair: Dr. Athan Zovoilis

5:05 – 5:50 pm	**Dr. Brett Trost, Sickkids** "Genomic architecture of autism from comprehensive whole-genome sequence annotation"	**1**
6:00 – 7:00 pm	**Supper**	
7:00 – 9:00 pm	**Cocktails and Networking**	

Wednesday, March 27th

7:00 – 9:00 am	**Breakfast – Vistas Dining Room (Badge requirement for entry)**	

The BioNet Outstanding Service Plenary Talk

Chair: Dr. Athan Zovoilis

9:00 – 9:25 am	**Dr. Steven Jones, Professor, FRSC, FCAHS** "Long reads, phasing and epigenomics"	**4**

Session 3: Health Omics 2

Chair: Dr. Georgia Balsevich

9:30 – 9:45 am	**Dr. William Hsiao, Simon Fraser University** **TBD**	**11**
9:50 – 10:05 am	**Dr. Mohamed Helmy, University of Saskatchewan** "Integrating Structural Bioinformatics and Artificial Intelligence in Designing High-Specificity Antibodies for COVID-19 Vaccine Development"	**12**
10:10 – 10:25 am	**Dr. Senthilkumar Kailasam, McGill University** "Ribosome footprint analysis workflow for studying neuronal stalled ribosome"	**13**
10:25 – 10:35 am	**David Enoma, University of Calgary** "Evolution-directed association studies (EDAS) in brain disorders give insight into the biology of Schizophrenia"	**23**
10:40 – 10:50 am	**Kevin Joannou, University of Alberta** "Thymic 4-1BB signals program the TCR αβ+ CD8αα intraepithelial lymphocyte population for survival"	**28**
10:50 – 11:20 pm	**Refreshment Break – Hallway outside Kinnear Centre 203**	

Session 4: Microbial Omics

Chair: Dr. Matthew Croxen

11:20 – 11:35 am	**Dr. Tarah Lynch, Alberta Precision Laboratories** "A modular bioinformatics system for public health genomics"	**14**
11:40 – 11:55 am	**Dr. Viraj Muthye, University of Calgary** "The highly repetitive genome of Myxobolus sp., a myxozoan parasite of fathead minnows"	**31**
12:00 – 12:10 pm	**Alyssa Butters, University of Calgary** "Colistin resistance and PmrB E123D and Y358N amino acid substitutions in Escherichia coli: A spurious association?"	**21**

Sponsor Talk: Illumina

Chair: Dr. Eric Merzetti

12:15 – 12:30 pm	**Sidki Bouslama, Illumina** "De novo bacterial assembly using Illumina long reads"	
12:30 – 1:30 pm	**Lunch – Hallway outside Kinnear Centre 203**	

Keynote Talk

Chair: Dr. Jennifer Geddes-McAlister

1:30 – 2:15 pm	**Dr. Jérôme Waldispühl, McGill University** "Leveling Up Citizen Science for Genomics"	**2**

Session 5: Agricultural Omics

Chair: Dr. Kirill Krivushin

2:25 – 2:40 pm	**Dr. Robert Gruninger, Agriculture and Agri-Food Canada** "Isolation and characterization of Bacteroides species associated with the formation of liver abscesses in feedlot cattle. "	**15**
2:45 – 2:55 pm	**Ryan Gourlie, Agriculture and Agri-Food Canada** "Replication of the virulence gene ToxB in the fungal plant pathogen Pyrenophora tritici-repentis"	**36**
3:00 – 3:30 pm	**Refreshment Break – Hallway outside Kinnear Centre 203**	

The BioNet 2024 Poster Session – Room KC 205

Judges: Dr. Elisabeth Richardson and Dr. Ryan Mercer

3:30 – 4:45 pm	Even Numbered Posters Present
4:45 – 6:00 pm	Odd Numbered Posters Present
6:00 pm	**End of Day 2: Attendees encouraged to group up and head into Banff for supper**

Thursday, March 28th

7:00 – 9:00 am **Breakfast – Vistas Dining Room (Badge requirement for entry)**

Plenary Talk

Chair: Travis Haight

9:00 – 9:25 am	**Dr. Michelle Scott, Université de Sherbrooke** "Using machine learning to identify the expression determinants of small nucleolar RNAs"	**5**

Session 6: Health Omics 3

Chair: Dr. Paul Gordon

9:30 – 9:45 am	**Dr. Jennifer Geddes-McAlister, University of Guelph** "Disruption of ClpX reverts fluconazole susceptibility for Cryptococcus neoformans through modified heme and ergosterol production"	**16**
9:50 – 10:05 am	**Dr. Rene Zahedi, University of Manitoba** TBD	**17**
10:10 – 10:25 am	**Dr. Igor Kovalchuk, University of Lethbridge** "Application of Drug Efficiency Index metric for analysis of post-traumatic stress disorder and treatment resistant depression gene expression profiles"	**18**
10:25 – 10:35 am	**Travis Haight, University of Lethbridge** "Nanopore Sequencing reveals a repetitive region in the Snhg14 locus that is connected with amyloid beta toxicity"	**26**
10:35 – 11:05 pm	**Refreshment Break – Hallway outside Kinnear Centre 203**	

Round-Table: Non-Academic Careers in STEM

Moderator: Dr. Athanasios Zovoilis

11:05 – 12:10 pm	**Panelists:** Sidki Bouslama - Illumina Athena Cochinamogulos – Genome Alberta Dr. Tarah Lynch – Alberta Precision Laboratories Dr. Ryan Mercer – Alberta Innovates Jim Slater – Cancer Care Manitoba
12:15 – 1:00 pm	**Closing Ceremonies – Awards / Final Comments**

4th Annual BioNet Conference – Poster Presentations

The 4th Annual BioNet Conference
Keynote and Plenary Speakers

Dr. Brett Trost, Sickkids

Genomic architecture of autism from comprehensive whole-genome sequence annotation

Autism is a neurodevelopmental condition characterized by social/communication difficulties, repetitive behaviors, and a restricted set of interests. Autism is highly heritable, making it critical to understand its genetic basis. In this talk, I will describe how we have used whole-genome sequencing data from large cohorts of autistic individuals and their family members to better understand how different types of rare genetic variation affect autism susceptibility, including single nucleotide variants, indels, structural variants, mitochondrial variants, and tandem repeat expansions. I will also discuss the substantial "missing heritability" in autism and how some of it may one day be explained using advances in bioinformatics methods.

Brett completed his PhD in Computer Science at the University of Saskatchewan, where he developed computational and statistical methods for studying phosphorylation-mediated cellular signaling. Subsequently, he was a Research Fellow at The Hospital for Sick Children (SickKids), where he used whole-genome sequencing data to study the genetics of autism and other neurodevelopmental and psychiatric conditions. Brett is the recipient of numerous awards and scholarships, including the Governor General's Academic Gold Medal, the Banting Post-Doctoral Fellowship, and the International Society of Psychiatric Genetics Richard Todd Award. Brett is currently a Scientist in the Molecular Medicine Program at SickKids, where he is interested in developing and applying machine learning methods for leveraging multi-omics data to improve our ability to identify and understand disease-associated genetic variation.

Dr. Jérôme Waldispühl, McGill University

Leveling Up Citizen Science for Genomics

Over the past decade, citizen science computer games have become a popular practice for engaging the public in research activities. This methodology had a noticeable impact in molecular and cell biology, where millions of online volunteers contributed to the classification and annotation of scientific data, but also to solve advanced optimization problems requiring human supervision. Yet, despite promising results, the deployment of citizen science initiatives through academic/professional web pages faces serious limitations. Indeed, the volume of human attention needed to process massive data sets and make state-of-the-art scientific contributions rapidly outpaces the participation and availability of online volunteers. To overcome this challenge, citizen science must transcend its "natural habitat" and reach out to the entire gaming communities. Therefore, one solution is to build partnerships with commercial video game companies that already assembled large communities of gamers.

In this talk, we describe how this approach can transform the impact of citizen science in genomics. We discuss our experience from Phylo, an online puzzle for gene alignment, to Borderlands Science, a massively multiplayer online game for microbiome data analysis. We show how to embeds citizen science tasks into a virtual universe to engage new user bases. These principles have profound implications for future citizen science initiatives seeking to meet the growing demands of biology.

Jérôme Waldispühl is an associate professor of Computer Science at McGill University. He holds a PhD from École Polytechnique (France), and previously was an instructor in Applied Mathematics at MIT (2006-2009). Jérôme conducts research in RNA structural bioinformatics and cheminformatics. He pioneered the use of video games to engage the public in genomic research with Phylo (2010), Colony B (2016), Borderlands Science (2020) and Project Discovery Phase 3 (2020), which he presented at the White House OSTP (2013), Québec Parliament (2016) and French Academy of Science (2018). He is a recipient of the Tomlinson Scientist Award and the Fessenden Professorship in Science Innovation.

Katrina Ingram, Ethically Aligned AI

Data and AI Ethics: A socio-technical perspective

Katrina Ingram is the Founder and CEO of Ethically Aligned AI, a company focused on helping organizations to drive better outcomes in the design, development, and deployment of AI systems. A seasoned executive, Katrina has over two decades of experience running both not for profit and corporate organizations in the technology and media sectors as well as experience in the public sector. She is a member of DAMA (data management professionals) and volunteers with several AI ethics organizations. She was named to the 100 Brilliant Women in AI Ethics list. Katrina holds an undergrad in business administration from Simon Fraser University, a Master of Arts in communications and technology from the University of Alberta and is an IAPP certified information privacy professional (CIPP/C). She combines her love of audio and interest in AI as the host of the podcast, AI4Society Dialogues. Katrina developed Canada's first micro-credential in AI Ethics in partnership with Athabasca University. She currently teaches at the University of Alberta and MacEwan University as a sessional instructor. She is a member of the Calgary Police Services Technology Ethics Committee and recently served as the City of Edmonton's Data Ethics Advisor.

Dr. Steven Jones, BC Cancer Research Institute and University of British Columbia

Long reads, phasing and epigenomics

Epigenomics, long-read sequencing, methylation, haplotyping, structural variation

Long-read nanopore sequencing has the potential to improve both the speed of genetic diagnosis as well as improving the diagnostic yield. The longer reads providing the ability to more accurately discover and define large scale structural variation. In addition, the alignment of the longer reads provide a more accurate assessment of regions that are highly repetitive and duplicated regions.

We have been utilizing long read sequencing to establish rapid turnaround genomic approaches to rapidly diagnose children with suspected genetic abnormalities in neonatal intensive care. We have also been using the epigenomic signals associated with the sequencing to help improve the rapidity of diagnosis through the confirmation of pathogenicity for variants initially determined as of unknown significance. The detection of phased methylation also allows the determination of imprinted regions. Coupled with long-range haplotyping the detection of imprinting this allows the assignment of the parental origin of all human autosomes providing the ability to significantly increase the speed and reduce the cost of hereditary cancer testing. The detection of other DNA modifications also increases the utility of long-read platforms for both diagnostics and advanced research.

Dr. Jones' research program is firmly entrenched in genome science to better understand the complete mutational landscape of cancers. His primary aim is to help uncover the diversity of genetic and genomic events that accrue to give rise to cancers, and which also encourage their evolution and maintain their progression. His laboratory extensively analyzes Next Generation genome and transcriptome data to achieve these goals. Dr. Jones has developed several novel computational approaches and methodologies to this end and has provided numerous insights into cancer dynamics, potential biomarkers, and therapeutic targets. A significant part of Dr. Jones research program relates to developing more precise cancer treatments by exploiting an individual's specific cancer genome profile. His research has identified numerous epigenetic targets that have the potential to be modulated in such a way as to reverse the effects of mutations within a cancer genome. Using computational approaches, his research team has identified and refined compounds that modify epigenetic programs in cancer. His laboratory also acts as a data analysis center for the Canadian Epigenetics, Environment and Health Research Consortium (CEEHRC).

Dr. Michelle Scott, Universite de Sherbrook

Using machine learning to identify the expression determinants of small nucleolar RNAs

Small nucleolar RNAs (snoRNAs) are non-coding RNAs well known for their role in ribosome biogenesis, many serving as guides for the site-specific modification of ribosomal RNA. In addition to this canonical function, the last two decades have demonstrated additional regulatory roles for subsets of snoRNAs in gene expression regulation the extent of which is still unknown. SnoRNAs display diverse expression strategies across eukaryotic species. In higher eukaryotes, many snoRNAs are embedded in the introns of longer genes referred to as their host genes while others are intergenic and expressed from their own promoters. Our recent transcriptomic profiling analyses of snoRNAs in human have revealed unexpected expression patterns including a majority of annotated snoRNAs never detected as expressed, intronic snoRNAs often poorly correlated with the abundance of their host gene and intronic snoRNAs within the same host gene displaying considerably different abundance values.

To better understand the determinants of snoRNA expression, we collected intrinsic and extrinsic features of human snoRNAs and their genomic context and used them to train machine learning models to predict the expression status of snoRNAs. Using Shapley Additive Explanations (SHAP values) to interpret the models' predictions, we determined that the most predictive features for snoRNA expression are conserved motifs, a stable global structure and strong terminal stem, and a transcribed locus and that these features are also predictive of expression in other vertebrates. We observe that these features explain well the varying abundance of snoRNAs embedded within the same host gene. The application of our predictor across several vertebrates revealed that only 1/3 of annotated snoRNAs are expressed per genome, as in human. Our results suggest that ancestral snoRNAs disseminated within vertebrate genomes, sometimes leading to the development of new functions and a probable gain in fitness, but many degenerating into pseudogenes, which should be taken into account in genomic annotations.

1. https://pubmed.ncbi.nlm.nih.gov/37072185/

Michelle completed an undergraduate degree at the Université de Montréal in Biochemistry and a Masters degree at the University of Calgary in Biochemistry and Molecular Biology followed by several semesters in computer engineering before discovering Bioinformatics. Following this revelation, she undertook a PhD in Bioinformatics at McGill University in Montreal under the co-supervision of Mike Hallett and David Thomas, studying the prediction and characterization of protein localization in the cell. She then moved to Geoff Barton's group at the University of Dundee in Scotland for her postdoc. Amongst her research interests during this time, she investigated the prediction of protein-protein interactions in human and the localization of proteins in the nucleolus. She also got initiated into the marvelous world of RNA (and particularly snoRNAs) by members of the Lamond group, working on snoRNAs regulating splicing and an evolutionary relationship between snoRNAs and miRNAs.

The 4th Annual BioNet Conference Speakers

Dr. Pinaki Bose, University of Calgary

Decoding MicroRNA Deregulation: Advancing Prognostication and Therapeutic Strategies in Oral Cancer.

Dr. Pinaki Bose completed his Ph.D. in basic cancer research with Dr. Karl Riabowol at the University of Calgary, investigating the role of the ING1 tumour suppressor protein in DNA damage signaling and apoptosis. After completing his Ph.D., Pinaki joined the Ohlson Research Initiative (ORI) as a postdoctoral fellow and trained in the molecular epidemiology of head and neck cancers under the supervision of Drs. Joseph Dort and Nigel Brockton. As part of a second postdoctoral fellowship, Pinaki trained at the BC Cancer Agency (BCCA) Genome Sciences Centre in cancer genomics and bioinformatics under one of the leading bioinformaticians in the world, Dr. Steven Jones. Pinaki was also a member of the personalized oncogenomics (POG) program at the BCCA. The POG initiative administers targeted therapies to recurrent/metastatic cancer patients based on their genomic profiles.

Dr. Bose currently directs the translational research program at the ORI, a multidisciplinary head and neck cancer research initiative working in close collaboration with surgeons, pathologists, bioinformaticians, statisticians and basic researchers. The Bose lab is investigating the biology of head and neck cancers with particular emphasis on mechanisms regulating lymph node metastasis (LNM). LNM is a potent predictor of worse prognosis in oral cancers, one of the most common cancers of the head and neck region. Towards this end, we are developing genome-wide screens in oral cancer cell lines and animal models to identify regulators of LNM.

The incorporation of automated phenotyping strategies to identify genes and pathways involved in sensory and metabolic hearing loss

Samah Ahmed[1], Kenneth I. Vaden Jr[2], Judy R. Dubno[2], Britt Drögemöller[1,3-5]

1 Department of Biochemistry and Medical Genetics, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, MB, Canada
2 Hearing Research Program, Department of Otolaryngology-Head and Neck Surgery, Medical University of South Carolina, SC, Charleston, USA
3 The Children's Hospital Foundation of Manitoba, Winnipeg, MB, Canada
4 CancerCare Manitoba Research Institute, Winnipeg, MB, Canada
5 Centre on Aging, University of Manitoba, Winnipeg, MB, Canada
Presenting author: britt.drogemoller@umanitoba.ca

Age-related hearing loss, Genome-wide association study, Automated phenotyping, ARHGEF28, KLHDC7B

Background: Age-related hearing loss (ARHL) affects one-third of the population over 65 years. The diverse pathology that underlies this heterogenous group of phenotypes likely involves distinct biological mechanisms. Although genome-wide association studies (GWAS) have uncovered genetic variants underlying self-reported ARHL, there are large gaps in our understanding of the precise genetic factors involved in specific hearing loss phenotypes. This may be due, in part, to challenges associated with accurate phenotyping for older adults with hearing loss. In this study, we used a mathematical model fitted to individual audiograms to estimate the magnitude of subtype components of ARHL.

Methods: We have obtained genomic and audiologic data from 26,622 healthy older individuals participating in the Canadian Longitudinal Study on Aging. By adopting an automated phenotyping approach, we derived metabolic and sensory estimates for each audiogram. We identified ears with better and worse hearing thresholds based on the calculated estimates. SNP-heritability testing was used to investigate the degree of variability that could be explained by genetics for each phenotype. GWAS was performed by linear regression and enrichment analyses were performed to identify biological pathways underlying the distinct hearing loss phenotypes.

Results: We found that metabolic estimates were higher for older compared to younger individuals and males showed more sensory hearing loss compared to females. For both metabolic and sensory estimates, ears with less severe hearing phenotypes showed higher heritability (h^2 = 0.064 and 0.106, respectively) and were included in downstream analyses. GWAS revealed that rs6453022, a missense variant in ARHGEF28, was significantly associated with the metabolic phenotype ($P=2.67\times10^{-9}$); while rs36062310, a missense variant in KLHDC7B, was significantly associated with the sensory phenotype ($P=2.37\times10^{-12}$). Enrichment analyses revealed differences in the biological pathways underlying the two hearing phenotypes, with the RhoA activity regulation pathway implicated in the metabolic phenotype, and pathways relating to sensory processing of sound by hair cells and the calcium/calmodulin signalling implicated in the sensory phenotype.

Conclusions: This is the first large-scale study to investigate the genetics of metabolic and sensory phenotypes separately. Our analyses have identified differences in the genetic variants associated with these two distinct hearing phenotypes. While it is known that hearing loss phenotypes involve diverse pathologies, this work represents the first large-scale study to uncover the genetic factors underlying these differences.

Advancing Precision Oncology: Machine Learning for Drug Response and Biomarker Discovery

Arvin Zaker[1,2], Arvind Mer[1,2,3]

1 Department of Biochemistry, Microbiology & Immunology, Faculty of Medicine, University of Ottawa
2 Ottawa Institute of Systems Biology
3 School of Electrical Engineering & Computer Science, University of Ottawa

Presenting author: *amer@uottawa.ca*

Biomarker, Machine learning, Pharmacogenomics, Personalized cancer medicine, Drug testing

In vivo drug testing using patient-derived xenografts (PDX) is an essential component of drug development for precision cancer medicine. They provide critical information for determining drug efficacy, toxicity, and biomarker discovery [1]. Currently, various biological endpoints, such as tumor volume, survival rate, growth rate, and pharmacokinetic parameters, are being used to assess drug efficacy in PDX [1,2]. However, the lack of a robust metric creates a hindrance in biomarker discovery, evaluating and comparing drug performances, and clinical decision-making [1,3]. Here, we present a machine learning strategy that integrates multiple biological endpoints into a single score for reliable drug response quantification. To achieve this, we first created xeMetron, an R package that computes 105 different response metrics from in vivo drug testing data. We applied this tool to our in-house dataset, which contains over 1,500 records of PDX-based drug testing experiments. To understand the relationship between various drug response metrics, a hierarchical clustering approach was used. We chose a subset of 10 response metrics whose performance was stable across a range of sample sizes. Using a multi-target regression technique, selected response metrics were combined to maximize the potential for genomic biomarker discovery. Using the combined metrics and PDX gene expression data, we trained a relevance vector machine (RVM) based model for drug response prediction. The model was validated in an independent in vitro [4] and clinical trial dataset, resulting in greater accuracy than alternative methods. These findings demonstrate that our method of combining multiple response metrics improves biomarker discovery and provides superior machine learning models for personalized cancer treatment.

1. https://pubmed.ncbi.nlm.nih.gov/31142512/
2. https://pubmed.ncbi.nlm.nih.gov/25185190/
3. https://pubmed.ncbi.nlm.nih.gov/28104906/
4. https://pubmed.ncbi.nlm.nih.gov/34850112/

Dr. Jason de Koning, University of Calgary

TBD

TBD

A scalable approach to fitting ZINB-WaVE models to single-cell gene expression data

Hamid Hamidi[1], Dinghao Wang[1], Quan Long[2], Matthew Greenberg[1]

1 Department of Mathematics and Statistics, University of Calgary, Calgary, Canada
2 Department of Biochemistry and Molecular Biology, University of Calgary, Calgary, Canada

Presenting author: *mgreenbe@ucalgary.ca*

Single cell, mixed model, pytorch, optimization, gene expression

We present a scalable, gradient-based procedure for fitting the flexible zero-inflated negative binomial "ZINB-WaVE" model of Risso et. al. (2018) to single-cell gene expression data. This model leads to a low-dimensional representation of the data accounting for its count nature, for zero-inflation (dropouts), and for overdispersion. We demonstrate that this model can outperform deep learning models at scale in terms of both natural loss functions and performance on various downstream tasks.

1. https://doi.org/10.1101/2023.07.29.5510622

Dr. William Hsaio, Simon Fraser University

TBD

TBD

Integrating Structural Bioinformatics and Artificial Intelligence in Designing High-Specificity Antibodies for COVID-19 Vaccine Development

Mohamed Helmy

Principal Scientist, Vaccine, and Infectious Disease Organization (VIDO), University of Saskatchewan, Saskatoon, Canada

Presenting author: *mohamed.helmy@usask.ca.com*

Structural bioinformatics, Vaccine development, COVID-19, Infectious diseases, Antibody design

The COVID-19 pandemic, caused by the novel SASR-CoV-2 virus, caused a global health crisis, necessitating urgent efforts for vaccine development. The emergence of diverse SARS-CoV-2 variants and the potential for future zoonotic outbreaks from the sarbecovirus subgenus highlight the need for continuous refinement of vaccine development strategies. Despite advancements, concerns persist regarding vaccine effectiveness against evolving variants. To address this, the ongoing project at the Vaccine and Infectious Diseases Organization (VIDO) at the University of Saskatchewan aims to design high-specificity coronavirus vaccines capable of broad coverage within the sarbecovirus subgenus, including SARS-CoV-2 variants [1]. The new vaccine contains several Receptor Binding Domains (RBDs) that binds to a wide range of sarbecovirus subgenus members aiming to widen the protection provided by this vaccine to include all know members and any future emerging variants from the same subgenus [2]. A key challenge in vaccine development is confirming the specificity of vaccine components, such as RBDs, which is crucial for regulatory approval. Utilizing specific monoclonal antibodies, we can validate the inclusion of each antigen subunit. However, identifying or designing antibodies with such high specificity is challenging. By employing a combination of structural bioinformatics and artificial intelligence (AI) [3], we have developed a method to optimize antibody design precision and efficacy. Our initial results are promising, emphasizing the need for further refinement to ensure biologically sound predictions. These refinements include accounting for various biological and chemical properties of the antibody and viral proteins, facilitating more comprehensive predictive models capable of addressing viral dynamics and variants. The designed antibodies undergo rigorous comparison with existing ones to identify candidates for repurposing or otherwise, novel development. By iteratively refining and validating our approach, we aim to significantly contribute to the development of highly specific and effective coronavirus vaccines, ultimately mitigating the impact of any future pandemic on global health.

1. *https://cepi.net/news_cepi/the-race-to-future-proof-coronavirus-vaccines/*
2. *https://www.researchsquare.com/article/rs-3656968/v1*
3. *https://pubmed.ncbi.nlm.nih.gov/37740287/*

Ribosome footprint analysis workflow for studying neuronal stalled ribosome.

Senthilkumar Kailasam [1], Jewel T-Y Li [2, 3], Francois Lefebvre [1,] Guillaume Bourque [1], Wayne Sossin [2,3]

1 Canadian Centre for Computational Genomics, McGill University, Montreal, Quebec H3A 0G1, Canada
2 Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, Montreal, Quebec H3A 2B4, Canada
3 Department of Anatomy and Cell Biology, McGill University, Montreal, Quebec H3A 0C7, Canada

Presenting author: senthil.duraikannukailasam@mcgill.ca

Neurons, ribosome, FMRP, riboseq, riboprofiling

The process of neuronal development and synaptic plasticity involves dynamic cellular events such as the transportation of ribosomes and local translation of mRNAs in distal sites. It provides a notable advantage of rapid supply of proteins when needed at synapses and axons, mediated by the reactivation of stalled polysomes. Previous work (1) had shown the enrichment of stalled ribosomes in RNA granules, emphasizing their role in local translation. However, the mechanisms leading to reversible ribosomal stalling in neurons remain unclear. In the present investigation, Riboprofiling was conducted on the pellet and polysome fractions of sucrose gradients generated from the brains of 5-day-old rats. A bioinformatic workflow was put together using existing software. This analysis revealed an increased presence of footprint reads originating from mRNAs that interact with FMRP. Additionally, there was an abundance of footprint reads derived from mRNAs related to cytoskeletal proteins implicated in neuronal development and increased ribosome occupancy on mRNAs encoding RNA-binding proteins. Compared to typical ribosome profiling studies, the footprint reads in this study was longer and mapped to reproducible peaks in mRNAs, enriched in motifs associated with mRNAs cross-linked to FMRP in vivo. These findings support a model in which specific mRNA sequences induce ribosomal stalling during translation elongation in neurons. Further, we explored the role of FMRP by comparing footprints from ribosomes in polysomes and RNA granules, extracted from FMR1 knockout and C57 mice. Through the analysis of protein contents, ribosomal footprint, and structural examination of RNA granules, the study seeks to provide insights into the specific contributions of FMRP in regulating neuronal translation and stalled polysome formation.

1. Anadolu, M. N., Sun, J., Kailasam, S., Chalkiadaki, K., Krimbacher, K., Li, J. T., ... & Sossin, W. S. (2023). *Journal of Neuroscience*, *43*(14), 2440-2459.

A modular bioinformatics system for public health genomics

Xiaoli Dong[1], Matthew Croxen[2,3], Andrew Lindsay[2], Stefan Gavriliuc[2], Vincent Li[2], Tarah Lynch[1,4]

1 Alberta ProvLab, Calgary, Canada
2 Alberta ProvLab, Edmonton, Canada
3 University of Alberta, Edmonton, Canada
4 University of Calgary, Calgary, Canada

Presenting author: *tarah.lynch@albertaprecisionlabs.ca*

Genomics, bioinformatics, public health, infectious diseases, microbiology, surveillance

In the realm of bioinformatics and computational biology, the utilization of pipeline workflow managers such as Nextflow or Snakemake has demonstrated significant efficacy. These tools enable the integration of multiple coding languages, facilitate reproducibility, and enhance accessibility for a broad user base.

While comprehensive bioinformatic pipelines create efficient methods for installing and running multi-step analyses, issues associated with adaptability can pose challenges for laboratories working with diverse organisms. Firstly, containerization ensures that a pipeline is both portable for sharing analytic methods and increases accessibility for laboratories lacking the resources or expertise to develop pipelines. However, this rigidity compromises flexibility in easily adding, removing, or updating individual components of the pipeline.

Secondly, the complexity of large, multi-step pipelines often necessitates bioinformatic expertise for maintenance, as the field rapidly evolves with development of new algorithms and tools. Although the operating system, current software, and dependencies are compatible at the time of pipeline development, updating any of these factors can disrupt processes. Evaluating and resolving these issues can require more time and resources than a laboratory has available, often resulting in pipelines becoming obsolete and necessitating alternative solutions.

The third issue involves redundancy across pipelines. Development following best practices incorporates quality control metrics and assurance operations for accurate and reliable input and output, which are present in all pipelines. Laboratories working on diverse targets can often have duplications of these quality processes along with other general tools. This redundancy becomes problematic when software tools need to be updated in multiple pipelines.

In this study, we propose a modularized approach to pipeline construction aimed at preserving adaptability and sustainability. In public health genomics, bioinformatic tools range from general applications to those tailored for identifying taxonomically specific genetic factors. By modularizing our bioinformatic pipelines, we can develop a system that is (1) resource-efficient by eliminating software duplication, (2) adaptable for diverse pathogens, and (3) resilient to component changes. We present findings from implementing a modularized analysis system within a public health laboratory setting, highlighting its utility across diverse pathogens.

Isolation and characterization of Bacteroides species associated with the formation of liver abscesses in feedlot cattle.

Gruninger RJ[1], Zaheer R[1], Chomistek N[1], King M[1], O'Hara EO[1], Ribeiro GO[2], Guan LL[3], Zovoilis A[4], McAllister TA[1]

[1] Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, Lethbridge, AB, Canada

[2] College of Agriculture and Bioresources, University of Saskatchewan, Saskatoon, SK, Canada

[3] Faculty of Agriculture, Life and Environmental Sciences, University of Alberta, Edmonton, AB, Canada

[4] Faculty of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, AB, Canada

Presenting author: robert.gruninger@agr.gc.ca

Beef Cattle, Bacteroides, Liver Abscess, Genomics, Metagenomics

Most feedlot-finishing diets in North America feed high-concentrate rations that are formulated to provide 10% or less dry matter as roughage. Diets that are low in forage increase the incidence of metabolic and digestive disorders and approximately 23-30% of cattle in feedlots develop liver abscesses. Severe liver abscesses are linked to reduced intake and decreased feed efficiency. In Canada, liver abscesses cost the industry ≈ $61.2 million annually. The biological processes involved in the development of liver abscesses is not well understood. Metataxonomic studies have shown that while *Fusobacterium necrophorum* is ubiquitous amongst these infections, close to 50% also have high levels of unknown Bacteroidetes. We present our recent efforts employing metagenomics and culturomics to examine the microbiology of liver abscesses in feedlot cattle. Shotgun metagenomics was conducted on liver abscesses collected from 12 cattle and MAGs were generated. *Fusobacterium* and *Bacteroides* were the most abundant MAGs, consistent with 16s based metataxonomic studies. Interestingly, a MAG was assembled from an unknown *Bacteroides*. Purulent material from the abscesses was subsequently used for culturomics experiments to isolate, identify, and characterize this species of *Bacteroides*. Phylogenomic analysis identified the Bacteroides isolates as *Bacteroides pyogenes* and a potentially novel species of *Bacteroides* related to *Bacteroides heparinolyticus,* consistent with the MAG analysis. The isolates encode several polysaccharide utilization loci operons involved in the metabolism of several carbohydrate sources and providing clues as to the role that they play in the development of liver abscesses. Work is ongoing to characterize the metabolism and antibiotic resistance of *Bacteroides* involved in the development of liver abscess. This work provides a substantial advancement in the understanding of liver abscess microbiology and is the first successful isolation and genomic characterization of the *Bacteroides* species that are found in bovine liver abscesses.

Disruption of ClpX reverts fluconazole susceptibility for Cryptococcus neoformans through modified heme and ergosterol production

Woods, Michael[1#], Bermas, Arianne[1#], Ball, Brianna[1], Muselius, Ben[1], Chan, Norris[1], Gutierrez-Gongora, Davier[1], Raouf-Alkadhimi, Fouad[1], McAlister, Jason[1], Shapiro, Rebecca S.[1], Sieber, Stephan A.[2], Geddes-McAlister, Jennifer[1]

[1]Molecular and Cellular Biology, University of Guelph, Guelph, Ontario, Canada, N1G 2W1
[2]Organic Chemistry, Technical University of Munich, Garching, Germany D-85748
[#]Authors contributed equally

*Presenting Author: jgeddesm@uoguelph.ca

Antifungal resistance, fluconazole, ClpX, Cryptococcus neoformans, quantitative proteomics

Fungal diseases impact the lives of millions of people across the globe. The opportunistic human fungal pathogen *Cryptococcus neoformans* causes cryptococcal meningitis in immunocompromised individuals with high fatality rates in response to limited treatment options. Moreover, the emergence of azole-resistant isolates in the clinic following prolonged treatment regimes, environmental fungicide exposure, and fungal evolution, threatens the outcome of current therapeutic options, endangering the survival of infected individuals. By quantitatively characterizing the proteomes of fluconazole-susceptible and -resistant *C. neoformans* strains using state-of-the-art tandem mass spectrometry combined with advanced bioinformatics platforms, we defined ClpX, an ATP-dependent unfoldase, as a target to overcome resistance. We discovered that disruption of ClpX through deletion or inhibition re-introduces fluconazole susceptibility into the resistant strains, rendering treatment effective once again. We further explored the mechanism of susceptibility and determined interruption to heme and ergosterol production associated with ClpX. Lastly, we are integrating *in silico* models of compound binding prediction to uncover novel strategies to specifically inhibit cryptococcal ClpX while limiting off-target effects against the host. Our results contribute to the understanding of novel mechanisms driving fluconazole resistance and provide support for targeting proteins as a therapeutic strategy to combat resistance.

Dr. Rene Zahedi, University of Manitoba

TBD

Application of Drug Efficiency Index metric for analysis of post-traumatic stress disorder and treatment resistant depression gene expression profiles

Nicolas Borisov[1], Yaroslav Ilnytskyy[2,3], Boseon Byeon[2,3,4], Olga Kovalchuk[2,3] and Igor Kovalchuk[2,3]

1 *Moscow Institute of Physics and Technology, 9 Institutsky lane, Dolgoprudny, Moscow Region, 141701, Russia; nicolasborissoff@gmail.com*
2 *Department of Biological Sciences, University of Lethbridge, Lethbridge, Alberta, T1K 3M4, Canada; sla-va.ilyntskyy@uleth.ca (Y.I.); bbyeon@gmail.com (B.B.); olga.kovalchuk@uleth.ca (O.K.)*
3 *Pathway Rx., 2 Fortress Rise SW, Alberta, T3H 4Z2, Canada*
4 *Biomedical and Health Informatics, Computer Science Department, State University of New York, 2 S Clinton St, Syracuse, NY 13202, USA*

Presenting author: *igor.kovalchuk@uleth.ca*

Post-traumatic stress disorder, treatment-resistant depression, next-generation sequencing data, gene expression, drug efficiency index, differentially expressed genes

Post-traumatic stress disorder (PTSD) is a severe mental illness with grave social, political, eco-nomic, and humanitarian implications. To apply the principles of personalized omics-based medicine to this psychiatric problem, we implemented our previously introduced drug efficiency index (DEI) to the PTSD gene expression data sets. Generally, omics-based personalized medicine evaluates individual drug action using two classes of data: (1) gene expression, mutation, Big Data pro-files, and (2) molecular pathway graphs that reflect the protein-protein interaction. In the particular case of the DEI metric, we evaluate the drug action according to the drug's ability to restore healthy (control) activation levels of molecular pathways. We have curated five PTSD and one TRD (treatment-resistant depression) cohorts of next-generation sequencing (NGS) and microarray hybridization (MH) gene expression profiles, which in total have 791 samples, including 379 cases, and 413 controls. To check the applicability of our DEI metrics, we have performed three differential studies with gene expression and pathway activation data: (1) cases samples vs control samples, (2) case samples after treatment or/and observation vs before treatment, and (3) samples from patients positively responding to the treatment vs those responding negatively, or non-responding ones. We found that the DEI values that use the signaling pathway impact activation (SPIA) metric were better than those that used the Oncobox pathway activation level (Oncobox PAL) approach. However, SPIA, Oncobox PAL, and DEI evaluations were reliable only if there were differential genes between case and control, treated and untreated samples.

Genetic association analysis of pediatric obsessive-compulsive behaviors, the role of imaging endophenotypes in mediating the relationship between genetic markers and quantitative symptom scores

Lilit Antonyan [1], S-M Shaheen [1], David R Rosenberg [2], Gregory L Hanna [3], Paul D Arnold [1]
1University of Calgary, Calgary AB, Canada
2Wayne State University, Detroit, MI, USA
3University of Michigan, Ann Arbor, MI, USA

Presenting author: lilit.antonyan1@ucalgary.ca

Psychiatry, Genetics, Quantitative, Endophenotype, Neuroimaging

Background: Obsessive-compulsive (OC) behaviors (OCB), characterized by intrusive thoughts and repetitive intentional behaviors are common in youth [1]. The cortico-striato-thalamo-cortical (CSTC) model is involved in the pathophysiology of OC traits [2]. Here, the goal is to determine the relationship between structural brain changes, genetic variation, and childhood OCB. We analyzed structural magnetic resonance imaging (MRI) as candidate neuroimaging endophenotypes for OCB.

Methods: MRI measures such as cortical thickness (CT), surface area (SA) and subcortical volume (SV) of specific brain regions were collected from 219 individuals from clinically diagnosed samples: 148 cases with mental disorders that manifest OC symptoms and 69 healthy controls. Polygenic risk score (PRS) analysis is applied to assess the probabilistic susceptibility for OCB using preselected phenotypes. The selection was based on the literature and on their correlation with OCB in our sample. The genotyped samples were run against the summary statistics of ~30,000 samples with matching phenotypes. PRS was conducted using (1) standard clumping+thresholding method: PRSice2 [3]; and (2) Bayesian shrinkage method: LDPred2-auto [4]. Mendelian Randomization analysis is applied to test the causality of the correlation between PRS and the endophenotypes [5].

Results: Two-step endophenotype prioritization yielded 8 CT, 6 SA and 5 SV estimates. PRS analysis on PRSice2 and LDPred2 showed that 2 brain regions (cortical thickness of superior parietal lobule, surface areas of lingual gyrus and inferior parietal lobule) have the best predictive ability with significant association results and may act as potential endophenotypes for OCB.
We will also identify the genetic risk markers that contribute to the high-risk scores.

Conclusion: This is the first study that includes child onset quantitative OC traits, genetic markers, and imaging measures in the same sample. Here we identify candidate neuroimaging endophenotypes that mediates effect of genetic markers on psychiatric OC score. These findings, if they remain significant in a larger sample would provide insight to the pathophysiology of OC traits, and the associated structural changes in specific brain regions.

1. https://doi.org/10.1097/00004583-199407000-00002
2. https://doi.org/10.1097/CHI.0b013e318185d2be
3. https://doi.org/10.1093/gigascience/giz082
4. https://doi.org/10.1093/bioinformatics/btaa1029
5. https://doi.org/10.1093/ije/dyg070

Power to detect episodic fitness shift at a small number of sites can be rescued with a covariate-informed branch-site type method

Mobina Kazemi Mehrabadi[1,*], Kabita Baral[1,*] and A.P. Jason de Koning, Ph.D.[1]

[1]Department of Biochemistry and Molecular Biology, University of Calgary, Calgary AB, Canada and Alberta Children's Hospital Research Institute (ACHRI), Calgary AB, Canada

*authors contributed equally

Presenting author: *Kabita.baral1@ucalgary.ca*

Evolution, Positive Selection, fitness shift, likelihood ratio, codon

Episodic positive selection (EPS) is generally believed to occur at a few sites and in a limited number of branches at a time, making it difficult to detect statistically. Branch-site tests are widely-used likelihood ratio tests that detect EPS along a prespecified lineage of a phylogeny at an unknown subset of sites. A limitation of power studies of branch-site methods has been a focus on detecting EPS simultaneously affecting an unrealistically large number of sites (e.g., >20%). We simulated sequence alignments under a set of fitness shift scenarios using a non-equilibrium time-heterogeneous mutation-selection codon model at a small number of sites (~1%). Then, we evaluated the power of different branch-site tests to detect signatures of episodic adaptation. Our results demonstrated that these tests have very low power to identify signatures of fitness shift at a small number of sites. We then implemented a covariate-based method to improve the identifiability of positive selection by adding site-specific data informing the probability of mixture class assignments to sites. This enhanced statistical power by a factor of 10 when covariate information was reliable (for the best model, power was 32.48%). When the results were filtered by the Kullback Leibler divergence (KLD) between ancestral and derived stationary distributions, we found that the covariate method can improve power drastically (up to 81% power for modestly large fitness shifts). Without covariate information, none of the traditional branch-site tests appear to have appreciable power, even in the presence of large fitness shifts simultaneously affecting multiple sites.

Colistin resistance and PmrB E123D and Y358N amino acid substitutions in Escherichia coli: A spurious association?

Alyssa Butters[a,b], Juan Jovel[a], Sheryl Gow[c,d], Karen Liljebjelke[a,b], Cheryl Waldner[d], Sylvia L. Checkley[a,b]

[a] Faculty of Veterinary Medicine, University of Calgary, Calgary, Alberta, Canada
[b] AMR--One Health Consortium, Calgary, Alberta, Canada
[c] Canadian Integrated Program for Antimicrobial Resistance Surveillance/FoodNet, Public Health Agency of Canada
[d] Department of Large Animal Clinical Sciences, Western College of Veterinary Medicine, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

Presenting author: alyssa.butters@ucalgary.ca

Antimicrobial resistance, colistin, PmrB, molecular epidemiology, Escherichia coli, phylogeny

Colistin is increasingly used as a last resort treatment for multi-drug and extensively drug resistant Gram-negative infections in humans. Although resistance mediated by plasmid-borne *m*obile *c*olistin *r*esistance (*mcr*) elements is well-recognized, chromosomally mediated colistin resistance is less understood. Amino acid substitutions in the PmrB sensor kinase have been associated with colistin resistance in *Escherichia coli*.

Illumina short-read whole genome sequencing was performed on 288 *E. coli* isolates from routine food animal feces, retail meat, and water surveillance. Phylogroup designations and genotypic antimicrobial resistance elements were determined *in silico* using ClermonTyping[1] and AMRFinderPlus[2] respectively.

Although all isolates were phenotypically non-resistant to colistin on agar screening antimicrobial susceptibility testing, 52.6% and 5.6% of isolates bore Y358N and E123D amino acid substitutions putatively associated with colistin resistance. Y358N-positive isolates were dominated by isolates from phylogroup B1 (97.4%), whereas all E123D-positive isolates belonged to phylogroup B2.

To further explore the associations between phylogeny, colistin susceptibility, and the PmrB Y358N/E123D amino acid variations, three additional *E. coli* datasets were assembled and similarly assessed for phylogroup and PmrB E123D/Y358N presence: 1) *E. coli* isolates from published studies that identified one or more isolates with either the Y358N or E123D amino acid variations and had publicly available whole-genome sequencing data (n=371), 2) *E. coli* isolates from published studies that were reported to be phenotypically colistin-susceptible by broth dilution antimicrobial susceptibility testing and had publicly available whole-genome sequencing submissions (n=880), and 3) 14,700 assemblies randomly sampled from all *E. coli* assemblies available in the National Center for Biotechnology Information public database[3].

In all datasets, significant differences in the distribution of both PmrB Y358N and E123D amino acid substitutions were found among phylogroups (Fisher's Exact test, $p < 0.001$). Within phylogroup B2 isolates, 94-100% demonstrated the PmrB E123D amino acid substitution and this variation was not found in any other phylogroup. The Y358N variation was found in ≥95% of phylogroup B1 and C isolates and rarely in other phylogroups. There was no evidence that colistin minimum inhibitory concentrations were higher among published colistin susceptible isolates that bore the Y358N or E123D amino acid

variations when compared to isolates not demonstrating these substitutions (Kruskal-Wallis test, $p > 0.05$).

Y358N and E123D amino acid substitutions are strongly associated with phylogroup in *E. coli* and neither alone appears sufficient to confer phenotypic colistin resistance. Reconsideration of previously identified associations between these variations and colistin resistance could be warranted. (396 words)

1. https://doi.org/10.1099/mgen.0.000192
2. https://www.nature.com/articles/s41598-021-91456-0
3. https://www.ncbi.nlm.nih.gov/assembly

Evolution-directed association studies (EDAS) in brain disorders give insight into the biology of Schizophrenia

David Enoma[1,2]

[1] Department of Biochemistry and Molecular Biology, Cumming School of Medicine, University of Calgary, Canada
[2] The Mathison Centre for Mental Health Research & Education, Hotchkiss Brain Institute, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Presenting author: *David.enoma@ucalgary.ca*

Evolutionary conservation, Genome-wide association studies, Schizophrenia, Kernel Machine, Genomics

Neuropsychiatric disorders such as schizophrenia (SCZ) present formidable challenges in diagnosis and treatment, mainly due to their multifaceted genetic origins. Schizophrenia has also been implicated as part of the Human Accelerated Regions that develop this human-specific brain disorder relative to our closest evolutionary relatives [1]. Evolution-directed association studies (EDAS) represent a promising methodology for unravelling the genetic underpinnings of SCZ by integrating evolutionary conservation data with genome-wide association study (GWAS) datasets. This study applied the EDAS framework to prioritize functionally relevant single nucleotide variations (SNVs) within a SCZ GWAS cohort, leveraging conservation scores derived from phastCons [2] and phyloP [3] methods. EDAS leverages genome evolutionary conservation scores between humans and 16 other primate species for SNVs from the UCSC genome browser [4].
Consequently, they are incorporated into a weighted kernel for association mapping via the Sequence Kernel Association Test (SKAT) [5]. Feature selection was conducted based on lower conservation scores (signifying Human acceleration), followed by feature aggregation to assess associations across European-American (EUR) and African American (AFR) ancestry populations. Application to SCZ GWAS data revealed significant associations, particularly in genes such as TENM4 and MAD1L1 among European ancestry populations [6] and REG4, CCDC25, ZNF395, CSMD1 and others among African ancestry populations [7]. Notably, these genes have been previously implicated in SCZ through GWAS and functional studies, thereby validating the effectiveness of the EDAS approach. Furthermore, pathway enrichment analysis uncovered pathways related to ubiquitin-mediated proteolysis, insulin resistance, and immune dysregulation, shedding light on potential mechanisms underlying SCZ pathogenesis. EDAS (17% EUR and 15% AFR) outperformed both mkTWAS (16% EUR and 14% AFR) [8] and unweighted SKAT (12% EUR and 14% AFR) based on validation rates (p-value <0.01) on SCZ Disgenet disease database [9]. These findings underscore the utility of EDAS in elucidating novel genetic associations and functional pathways related to neuropsychiatric disorders.
In conclusion, our study demonstrates the efficacy of the EDAS approach in integrating evolutionary conservation data to prioritize SNVs and uncover genetic associations in SCZ. These findings contribute to our understanding of the complex genetic architecture of neuropsychiatric disorders and may inform future diagnostic and therapeutic strategies. This study emphasizes the importance of leveraging evolutionary conservation information in advancing our understanding of SCZ and other Brain disorders, paving the way for personalized medicine approaches tailored to individual patients' genetic profiles.

[1] M. J. Hubisz and K. S. Pollard, “Exploring the genesis and functions of Human Accelerated Regions sheds light on their role in human evolution,” *Current Opinion in Genetics & Development*, vol. 29, pp. 15–21, Dec. 2014, doi: 10.1016/j.gde.2014.07.005.

[2] A. Siepel and D. Haussler, “Phylogenetic Estimation of Context-Dependent Substitution Rates by Maximum Likelihood,” *Molecular Biology and Evolution*, vol. 21, no. 3, pp. 468–488, Mar. 2004, doi: 10.1093/molbev/msh039.

[3] K. S. Pollard, M. J. Hubisz, K. R. Rosenbloom, and A. Siepel, “Detection of nonneutral substitution rates on mammalian phylogenies,” *Genome Res.*, vol. 20, no. 1, pp. 110–121, Jan. 2010, doi: 10.1101/gr.097857.109.

[4] M. Blanchette *et al.*, “Aligning Multiple Genomic Sequences With the Threaded Blockset Aligner,” *Genome Res.*, vol. 14, no. 4, pp. 708–715, Jan. 2004, doi: 10.1101/gr.1933104.

[5] M. C. Wu *et al.*, “Powerful SNP-Set Analysis for Case-Control Genome-wide Association Studies,” *The American Journal of Human Genetics*, vol. 86, no. 6, pp. 929–942, Jun. 2010, doi: 10.1016/j.ajhg.2010.05.002.

[6] Cross-Disorder Group of the Psychiatric Genomics Consortium, “Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis,” *The Lancet*, vol. 381, no. 9875, pp. 1371–1379, Apr. 2013, doi: 10.1016/S0140-6736(12)62129-1.

[7] G. Donohoe *et al.*, “Neuropsychological effects of the CSMD1 genome-wide associated schizophrenia risk variant rs10503253,” *Genes, Brain and Behavior*, vol. 12, no. 2, pp. 203–209, 2013, doi: 10.1111/gbb.12016.

[8] C. Cao *et al.*, “Disentangling genetic feature selection and aggregation in transcriptome-wide association studies,” *Genetics*, vol. 220, no. 2, p. iyab216, Feb. 2022, doi: 10.1093/genetics/iyab216.

[9] J. Piñero, J. Saüch, F. Sanz, and L. I. Furlong, “The DisGeNET cytoscape app: Exploring and visualizing disease genomics data,” *Computational and Structural Biotechnology Journal*, vol. 19, pp. 2960–2967, Jan. 2021, doi: 10.1016/j.csbj.2021.05.015.

Democratisation of snoGloBe and evaluation of its predictive capacity in eukaryotes

Morgane Govone[1], Gabrielle Deschamps-Francoeur[1], Michelle S. Scott[1]

1 UniversitédeSherbrooke, Sherbrooke,Québec,Canada

Presenting author: *Morgane.Govone@USherbrooke.ca*

snoRNA, Machine Learning, Predictor, Web Server, RNA-RNA Interactions, high-throughputRNA sequencing

Box C/D snoRNAs are non-coding RNAs known for their role in ribosome biogenesis across all eukaryotes, many serving as guides for the modification of ribosomal RNA (rRNA). However, recent research in mammals reveals additional roles in gene expression regulation, such as the modulation of alternative splicing and messenger RNA (mRNA) transcripts stability.
With the objective of investigating the scope of snoRNA functions on a large scale by studying their targets, our group created SnoGloBe [1], a predictor of C/D snoRNA interactions built and validated in human. SnoGloBe is a command line tool that employs Machine Learning to identify snoRNA binding sites in various RNA targets. We aim to test its predictive capacity in different eukaryotes and improve the usability of snoGloBe by creating a webserver.
To determine the range of eukaryotic organisms in which snoGloBe is predictive, we compared its transcriptome-wide predictions to RNA-RNA interactions measured using high-throughput sequencing datasets, generated in yeast and mouse. These sequencing datasets were all generated using crosslinking, proximity ligation and RNA sequencing with [2,3] or without an initial protein pulldown [4,5]. We analyzed these data de novo to identify interactions involving at least one snoRNA, finding 3898 distinct mRNAs and 275 snoRNAs in mouse and 1616 distinct mRNAs, 41 tRNAs and 76 snoRNAs in yeast. The comparison with snoGloBe predictions revealed that it detects between 25 and 50% of interactions detected by sequencing using the most stringent thresholds.
SnoGloBe is currently a bioinformatics tool that can only be used on the command line. Thus in parallel with our above research, we are creating a web server that will make it easier to use for the wide scientific community. Users will simply be required to provide the inputs, i.e. the sequences of our snoRNAs of interest, and information about the target(s) we want to test (sequence or id, location, and biotype). Optionally, the user can also control different parameters including the minimum score threshold, the minimum number of positive consecutive windows and the steps between the windows, without having to write any command lines. The webserver input and output will be presented.
Overall, the study increases the user friendliness of snoGloBe, investigates its organismal range and compares the distribution of snoRNA targets across eukaryotes widening our understanding of snoRNA biology.

1. https://doi.org/10.1093/nar/gkac475
2. https://doi.org/10.1101/2021.07.22.451324
3. https://doi.org/10.12688/wellcomeopenres.14735.2
4. https://doi.org/10.1016/j.cell.2016.04.028
5. https://www.nature.com/articles/ncomms12023

Nanopore Sequencing reveals a repetitive region in the Snhg14 locus that is connected with amyloid beta toxicity

Travis Haight[1,2], Riya Roy[1,2], Sani Zaidi[2], Luke Saville[2], Igor Kovalchuk[1], Athanasios Zovoilis[2]

1: Department of Chemistry and Biochemistry, University of Lethbridge
2: Department of Medical Genomics, Cancer Care Manitoba
Presenting author: *haight@uleth.ca ; travis.haight@umanitoba.ca*

Nanopore, Amyloid-beta Toxicity, long-ncRNA, circRNA, Snhg14

Long non-coding RNAs (lncRNAs) have emerged as critical regulators of gene expression, with a direct impact on biological processes and disease states. Among this class of RNA, the small nucleolar RNA host gene 14 (Snhg14) locus has attracted attention for its potential role in various cellular functions and pathological disorders such as Prader-Willi Syndrome and Angelman syndrome [1]. Despite its recent attention, the landscape governing Snhg14 locus's regulation, expression and functional implications remains poorly understood due to its repetitive nature. In this study we utilize Ilumina and Nanopore Next generation sequencing to shine light on this very important locus as a response to Amyloid-Beta Toxicity.

1. https://doi.org/10.3390/genes14010097

Genomic Interactions between Indacaterol and Carbachol in Human Airway Epithelial Cells: Implications for the Treatment of Obstructive Lung Diseases

Varuna Jayasinghe[1], Radhika Joshi[1], Taruna Joshi[1] & Mark A. Giembycz[1]

[1] Lung Health Research Group, Department of Physiology & Pharmacology, Snyder Institute for Chronic Diseases, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada

Presenting author: *varuna.jayasinghe@ucalgary.ca*

COPD, airway epithelial cells, gene expression, acetylcholine, long-acting B2-adrenoceptor agonists, long-acting muscarinic receptor antagonists, functional annotation

Susceptibility to recurrent exacerbations is a cardinal feature of chronic obstructive pulmonary disease (COPD) and increases the risk of premature mortality. Clinical trials have shown that long-acting muscarinic receptor antagonists (LAMAs) protect against exacerbations possibly by suppressing airway inflammation. Indeed, there is evidence that acetylcholine (ACh) released by cigarette smoke is pro-inflammatory in the airways. Typically, inflammation occurs secondarily to changes in gene expression. In this study, we have explored the pro-inflammatory potential of carbachol (CCh), a hydrolysis-resistant analogue of ACh, by measuring gene expression changes in the BEAS-2B human airway epithelial cell line. This exercise was performed in the absence and presence of a β2-adrenoceptor agonist, indacaterol (Ind), which is typically given to COPD patients as a bronchodilator but, paradoxically, can up-regulate genes that may be detrimental to lung health.

Next-generation mRNA sequencing was employed to detect gene expression changes in BEAS-2B cells over a period of 18 hours. CCh, Ind and Ind+CCh significantly ($q \leq 0.05$) regulated the expression (induced ≥1.5-fold; repressed ≤0.67-fold) of 20 (20 induced; 0 repressed), 869 (624 induced, 256 repressed) and 1027 (691 induced, 352 repressed) unique genes, respectively. Although CCh was a weak transcriptional activator per se, it significantly ($P \leq 0.05$) modulated the expression of 143 (22.9%) unique Ind-induced genes and 17 (6.6%) unique Ind-repressed genes at one or more time-points in either a positive or negative manner. The expression of 105 Ind-induced mRNAs was augmented by CCh in either a supra- (39) or infra-additive manner (66), including the pro-inflammatory genes IL6, NR4A3, JUNB, CCL2 and CITED4. Using the database for annotation, visualization and integrated discovery (DAVID), functional annotation of the genes regulated by Ind, CCh and Ind+CCh captured 104 enriched gene ontology terms/KEGG pathways and 12 functional clusters with transcriptional control (enrichment score = 10.32) being the most enriched theme. The regulation of cytokine/chemokine release (i.e., inflammation) was also significantly enriched (ES = 2.51). Manual curation of these enriched terms defined six, broad functional categories; “Transcription”, “Signalling”, “Development, differentiation, proliferation and apoptosis”, “Inflammation” and “Chemotaxis and migration”. In all categories, the gene count and associated significance were greater with Ind+CCh relative to Ind alone. We conclude that the ability of CCh to enhance the transcriptomic and functional signature of Ind implies that these two stimuli can interact at a genomic level, which may have implications for the treatment of COPD.

1. https://doi.org/10.1164/rccm.201609-1794CI
2. https://doi.org/10.2147/COPD.S285867
3. https://doi.org/10.1016/j.pharmthera.2007.05.007
4. https://doi.org/10.1124/jpet.118.249292

Thymic 4-1BB signals program the TCR αβ+ CD8αα intraepithelial lymphocyte population for survival

Kevin Joannou[1], Dominic Golec[2], Julia May[1], Timothy Starr[3], Troy Baldwin[1]

1University of Alberta, Edmonton, Alberta, Canada, 2NIAID, NIH, Washington, DC, United States, 3University of Minnesota, Minnesota, United States.

Presenting Author: joannou@ualberta.ca

4-1BB, Agonist Selection, CD8αα Intraepithelial Lymphocytes, Single-Cell RNA Sequencing, T Cell Development

Introduction: T Cell Receptor (TCR) αβ+ CD8αα Intraepithelial Lymphocytes (IEL) are a population of αβ T cells that reside between intestinal epithelial cells and perform regulatory roles. While most developing αβ T cells that experience agonist signals in the thymus are deleted as a central tolerance mechanism, precursors of TCRαβ+ CD8αα IEL (IELp) require agonist TCR signals for selection. The necessary intrathymic signals that allow these agonist-selected cells to be diverted away from clonal deletion in the thymus to unique tissue resident roles in the periphery remains unclear.

Methods/Results: We completed a trajectory analysis of thymic IELp using scRNA-seq and flow cytometry and identified 4-1BB, a costimulatory receptor in the Tumor Necrosis Factor Receptor superfamily, as a candidate for contributing to IELp maturation. 4-1BB expression on IELp correlated with selection of IELp away from apoptosis and towards maturity. Flow cytometry analysis suggests that 4-1BB is transiently upregulated during thymic IELp development, and experiments with Nur77-GFP reporter mice indicated that 4-1BB is expressed on the most highly self-reactive IELp. Analysis of 4-1BB-deficient mice and competitive mixed bone marrow chimeras supported a role for 4-1BB in generating CD8αα IEL but not in completing the thymic stages of IELp maturation, nor in upregulation of trafficking molecules needed to enter the intestinal epithelium. However, experiments culturing IEL in-vitro with IL-15 suggest a survival defect in 4-1BB KO IEL.

Conclusions: Taken together, our data suggest that the most highly TCR signaled IELp undergo a short window of 4-1BB signaling in the thymus that programs them for long-term survival as tissue residents. Significantly, this may have implications for the programming of innate-like tissue resident T cells more broadly.

Funding Sources: CIHR, LKSIoV, FOMD

Comparative Genomics of Lipid Droplet-Associated Proteins Across Eukaryotic Diversity

Emily M. Kingdon[1], Sarah M. Orton[1], Elisabeth Richardson[1]

1 Mount Royal University, Calgary, Alberta

Presenting author: eking057@mtroyal.ca

Genomics, Homology Searching, Bioinformatics, Lipid Droplet, Eukaryotic Diversity

Lipid droplets (LDs) are a hydrocarbon storage organelle represented across biologically diverse organisms (Lundquist et al., 2020). They are present across all forms of life and are essential for lipid storage and metabolism. Though these organelles are widely distributed, very little is known about how LDs emerged, evolved, or whether extant LDs share a common origin or organellular ancestor. It has also been shown that LDs facilitate an impressive variety of cellular functions, including stress responses, cellular signaling, membrane remodeling, and ROS- scavenging, and the distribution of these functions across eukaryotic diversity is unknown. (Lundquist et al., 2020). We have used comparative genomics to examine the evolution of proteins associated with LD biogenesis across eukaryotic diversity, to ultimately determine if LDs in different eukaryotic supergroups are homologous. We used reciprocal BLAST strategies to identify divergent homologues of lipid droplet related proteins across eukaryotic diversity (Hirst et al., 2014). Using representative organisms from every group in the eukaryotic tree as defined in Burki et al. (2020), we used this strategy to find homologs of LD related proteins in the genomes of the representative organisms. By studying the LD-associated protein SEIPIN, we were able to characterize a previously unidentified pan-eukaryotic distribution, as well as considerable paralogue expansion within mammals.

1. https://pubmed.ncbi.nlm.nih.gov/32348789/
2. https://elifesciences.org/articles/2866
3. https://pubmed.ncbi.nlm.nih.gov/31606140/

Experimental Design Factors Controlling Power to Detect Episodic Fitness Shifts using Comparative Genomic Sequencing Data

Oliver A.S. Lyon 1, A.P. Jason de Koning 1

1 University of Calgary, Calgary, Canada

Presenting author: oliver.lyon@ucalgary.ca

A central challenge in comparative genomics is identifying the genomic basis for lineage-specific adaptations to changing functional requirements. One approach to addressing this problem using genomic sequence data alone, which overcomes limitations of traditional dN/dS approaches, is to model site-heterogeneous sequence-fitness relationships and to infer changes to such relationships along the branches of a phylogeny (inference of 'fitness shifts'). This requires very parameter-rich models and correspondingly large datasets. However, the factors shaping the detectability of fitness shifts, as well as the distinguishability of fitness shifts from other time-heterogeneous forces such as variation in effective population sizes across lineages, are unclear. Here, we develop a framework for identifying such factors by measuring the asymptotic distinguishability of models of sequence evolution along a fixed phylogeny. We developed an efficient C++ library for modelling and inference of time-heterogeneous (Markov-modulated) mutation-selection codon substitution models – a class of models with an explicit population genetics basis. Using this framework, we measured distinguishability of fitness-shift from time-homogeneous evolution, and fitness-shift from changes in the effective population size in terms of the Kullback-Leibler divergence between models. Using these measurements, we show how asymptotic power analysis can be easily performed to assess minimum sample sizes needed to achieve reasonable power levels. Notably, we focused our analysis on the SARS-CoV-2 phylogeny, given the pressing need to identify functional shifts among variants of concern amidst the ongoing pandemic.

The highly repetitive genome of *Myxobolus* sp., a myxozoan parasite of fathead minnows

Viraj R. Muthye[1,2], Aralia Leon Coria[2,3], Hongrui Liu[2,3], Constance A. M. Finney[2,3], Cameron P. Goater[2,4], James D. Wasmuth[1,2]*

[1.] Faculty of Veterinary Medicine, University of Calgary, Calgary, Canada
[2]. Host-Parasite Interactions Research Training Network, University of Calgary, Calgary, Canada
[3]. Department of Biological Sciences, University of Calgary, Calgary, Canada
[4]. Department of Biological Sciences, University of Lethbridge, Lethbridge, Canada

Presenting author: virajrajendra.muthye@ucalgary.ca

myxozoa, parasite, genome, transposons, nanopore, cnidaria

Myxozoans are an enigmatic group of around 2,400 parasites within Cnidaria, a phylum that also includes jellyfish, corals, hydra, and sea anemones. All myxozoans are greatly reduced in size and morphology compared to the free-living members of this phylum. In their life cycle, myxozoans alternate between invertebrates (annelids or bryozoans) and vertebrates (mostly fish). They are best known for causing disease in economically important and iconic fish across the world. For example, *Myxobolus cerebralis* causes Whirling Disease, which can kill 90% of infected juvenile salmonid fish. Despite the economic, ecological, and conservation challenges posed by myxozoans, there are no effective drugs to treat myxozoan infections. Furthermore, the genomes of only eight of the ~2,400 myxozoan species have been sequenced.

Myxozoan diversity is very likely largely underestimated and new species are being discovered each year. In 2017, a potentially new myxozoan species was identified in Alberta. *Myxobolus* sp. causes distinct lesions in fathead minnows, which are ultimately fatal. Here, we sequenced, assembled, and analyzed the genome of *Myxobolus* sp. to understand how the parasite interacts with its fish host and to identify potential strategies to counter this emerging threat.

We used the Oxford Nanopore long-read technology for sequencing the genome of this parasite. At 185 Mb, the *Myxobolus* sp. genome is the largest myxozoan genome sequenced so far. This large genome size is, in part, due to the high repetitive content; 68% of the genome was comprised of interspersed repeats, with the MULE-MuDR transposon covering 18% of the *Myxobolus* sp. genome. Like other myxozoan genomes, the *Myxobolus* sp. genome has lost many genes well conserved in other eukaryotes. However, we also identified multiple expansions in gene families (serine proteases, hexokinases, and FLYWCH domain-containing proteins) which suggests their functional importance in the parasite. The mitochondrial genome of *Myxobolus* sp. displayed several intriguing features different from the typical animal mitochondrial genome: 1) it encodes only five of the thirteen protein-coding genes typically found in animals, 2) the *atp6* gene has been transferred to the nucleus and has acquired a mitochondria-targeting signal, and 3) the *rnl* gene is fragmented.

Our study provides valuable insights into myxozoan biology and identifies promising avenues for future research. We also propose that *Myxobolus* sp. is a promising myxozoan model to explore host-parasite interactions in these parasites.

Multiomic analysis to identify host and microbiome contributions to digestibility in beef cattle

E. O'Hara[1]; M. M. Dubois[2]; N. P. Payne[2]; G. O. Ribeiro[2]; R. J. Gruninger[1]

1 Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, Lethbridge, Canada.
2 Department of Animal and Poultry Science, University of Saskatchewan, Saskatoon, Canada.

Presenting author: eoin.ohara@agr.gc.ca

Microbiome, metagenome, transcriptome, rumen, feed efficiency

This study evaluated beef heifers selected for high (efficient) or low (inefficient) digestible fibre intake (DFI). Initial analysis showed that high DFI animals had reduced methane production versus low-DFI under a high forage diet. Using the same cohort of animals maintained on a further 4 diets of varying forage: concentrate ratios, we employed multi-kingdom amplicon sequencing and metagenome shotgun sequencing of rumen digesta and feces alongside RNA sequencing of rumen epimural samples to evaluate the compositional and functional interplay between different microbial groups, and their relationship with host gene expression in cattle divergent for DFI. Samples were collected from 16 cattle during 2 metabolism trials, comprising 5 diets. Amplicon sequencing analysis was conducted using QIIME2; 16S rRNA and 18S rRNA reads were analysed using the SILVA database, while LSU (fungal) sequences were analysed using a custom D1/D2 database. Additional analysis of archaeal 16S rRNA sequences was conducted using the Rumen and Intestinal Methanogens (RIM) database. Metagenome shotgun reads underwent a two-pass classification with Kraken2 using a database of prokaryotic genomes derived from the GTDB taxonomy, with the unclassified output undergoing classification using a custom database containing all NCBI protozoa, fungi, and phage genomes, enriched with selected rumen-specific ciliate and fungal genomes. Downstream analysis of taxonomic data from all microbiome work was conducted in R, and differentially abundant taxa were identified using ANCOM-BC and Aldex2. Functional analysis of metagenome contigs using the CAZY database implemented in dbCAN3 is ongoing. RNA-seq data was analysed using the ARS-UCD reference genome, with identification of DE genes conducted using DeSeq2. Preliminary results indicate no major effect of DFI ranking on host gene expression, bacterial 16S rRNA, or metagenome compositional profiles. Several bacterial genera were differentially abundant between digestibility groups ($P<0.05$), but these were all minor (<0.01%) members of the microbiome. Fungal and methanogen communities different significantly ($P<0.05$) according to DFI group, with efficient (high DFI) containing and more diverse communities under high-grain diet ($P<0.05$). The same difference showed a tendency toward significance for the 18S rRNA protozoa data ($P<0.1$). These preliminary data indicate that the microbial factors underpinning divergence in efficiency measured by DFI vary according to diet and may be more prominent in the non-bacterial fraction of the microbiome. Ongoing functional analysis of metagenome data as well as integration of multiomic data will provide deeper insight into these relationships and how they contribute to feed digestibility and efficiency in cattle.

Deconvolution of immune cell infiltration in solid tumours from DNA alignment signatures

Haley Pederson[1], Paul Gordon[1]

1 University of Calgary, CSM Centre for Health Genomics and Informatics, Calgary, Canada

Presenting author: haley.pedersen@ucalgary.ca

Cancer, tumour-infiltrating lymphocytes, VDJ recombination, gapped alignment, whole-genome sequencing

Solid tumour micro-environments contain a diverse mixture of cell populations in addition to malignant cells. Tumour-infiltrating lymphocytes (TILs), which characterize part of the host immune response, are a key prognostic feature in many cancers and help determine which patients may be candidates for immune-based therapies [1], [2]. Current methods to detect TILs include H&E, flow cytometry, traditional and multiplexed fluorescent immunohistochemistry, and more recently, deconvolution of gene expression profiles from bulk tumour transcriptomes [3]. However, these methods can be subjective, require resource allocation to immune-specific workflows, and in the case of transcriptome profiling, rely on reference expression profiles that may not generalize to dysregulated cells within tumour micro-environments [4], [5].

Variable–diversity–joining (VDJ) rearrangement is a type of somatic recombination that occurs exclusively in developing lymphocytes [6]. When mapped to a reference genome, these rearrangements result in complex, gapped alignments of lymphocyte-derived DNA to VDJ regions. We posit that by examining coverage and fragment length signatures of VDJ mapped reads, it is possible to detect the presence and proportion of lymphocytic cells within admixed tumour samples directly from genome sequencing data. Thus, to increase the accessibility and generalizability of tumour immune profiling, we are developing the first DNA alignment-based TIL detection approach: *VDJump.* We plan for *VDJump* to integrate seamlessly into popular matched tumor-normal whole-genome sequencing pipelines and allow retrospective analysis of existing bulk sequencing data. *VDJump*'s performance will be assessed in 400 matched tumour-normal samples of various cancer types with accompanying clinical and histological data.

1. https://www.ncbi.nlm.nih.gov/pmc/arBcles/PMC3137407/
2. https://pubmed.ncbi.nlm.nih.gov/33139907/
3. https://www.ncbi.nlm.nih.gov/pmc/arBcles/PMC5564448/
4. https://www.ncbi.nlm.nih.gov/pmc/arBcles/PMC4739640/
5. https://www.ncbi.nlm.nih.gov/pmc/arBcles/PMC4852857/
6. https://www.nature.com/arBcles/302575a0

Elucidating the role of Adenosine-to-Inosine editing in SINE RNAs: Connecting the dots between editing and cellular response to stress

Riya Roy[1,2], Travis Haight[1,2], Luke Saville[1,2], Liam Mitchell[1,2], Athanasios Zovoilis[1]

Affiliations:
1. Southern Alberta Genome Sciences Centre, University of Lethbridge, Alberta
2. Department of Biochemistry and Medical Genetics, University of Manitoba

Presenting author: riya.roy@uleth.ca

Non-coding RNAs, SINEs, RNA editing, Murine B2, Cellular stress

The mammalian genome contains only ~2% of protein coding region, raising the questions about the significance of the rest 98% noncoding region. Short Interspersed Nuclear Elements (SINEs) are one of the most abundant classes of non-coding RNAs involving murine B2 and human Alu as the most frequent ones[1]. Adenosine-to-Inosine RNA editing by adenosine deaminase acting on RNA (ADAR) enzymes, is a prevalent post-transcriptional modification that predominantly targets double stranded RNA. ADARs majorly modifies Alu and B2 SINEs[3]. Various studies have revealed the significant involvement of SINE RNAs in cellular response to stress[2]. On the other hand, the A-to-I RNA editing has been reported to be connected with the pathogenesis of several neurological diseases. However, the role of A-to-I editing in SINE RNA stability remains unclear. Here, we are trying to investigate the connection between RNA editing and the stability of ADARs' major targets-B2 SINE RNAs, focusing more on the correlation between SINE RNA processing ratio and A-to-I editing during response to cellular stress in mouse.

1. Yakovchuk P, Goodrich JA, Kugel JF. B2 RNA and Alu RNA repress transcription by disrupting contacts between RNA polymerase II and promoter DNA within assembled complexes. Proc Natl Acad Sci U S A. 2009 Apr 7;106(14):5569-74. doi: 10.1073/pnas.0810738106. Epub 2009 Mar 23. PMID: 19307572; PMCID: PMC2667051.
2. Cheng Y, Saville L, Gollen B, Isaac C, Belay A, Mehla J, Patel K, Thakor N, Mohajerani MH, Zovoilis A. Increased processing of SINE B2 ncRNAs unveils a novel type of transcriptome deregulation in amyloid beta neuropathology. Elife. 2020 Nov 16;9:e61265. doi: 10.7554/eLife.61265. PMID: 33191914; PMCID: PMC7717908.
3. Huntley MA, Lou M, Goldstein LD, Lawrence M, Dijkgraaf GJ, Kaminker JS, Gentleman R. Complex regulation of ADAR-mediated RNA-editing across tissues. BMC Genomics. 2016 Jan 15;17:61. doi: 10.1186/s12864-015-2291-9. PMID: 26768488; PMCID: PMC4714477.

An integrative transcriptomics approach challenges the completeness of human snoRNA annotations.

Alphonse Birane Thiaw[1], Michel Renaud[1], Étienne Fafard-Couture[1], Danny Bergeron[1], Sonia Couture[1], Sherif Abou Elela[1], Michelle Scott[1]

1 Université de Sherbrooke, Sherbrooke, Canada (Quebec)

Presenting author: alphonse.birane.thiaw@usherbrooke.ca

snoRNA, TGIRT-seq, Stringtie, Parclip, eclip, Dyskerin depletion, NOP58 depletion

Small nucleolar RNAs (snoRNAs) are non-coding RNAs present in all eukaryotes and best known for their involvement in ribosomes biogenesis. In mammals, most snoRNAs are located in the introns of longer genes, however some are expressed from intergenic regions. Based on their motifs and functions, snoRNAs are classified into two groups: box H/ACA and box C/D snoRNAs which guide respectively pseudouridylation and 2'-O-ribose methylation of their targets, many of which are in ribosomal RNA. Besides these canonical functions, some snoRNAs are also known for their involvement in gene expression regulation at several levels. Despite the breadth of functionality described for snoRNAs, the majority are poorly characterized, and recent transcriptomic studies have both identified unannotated human snoRNAs missing from current references and shown that many annotated human snoRNAs are not expressed. We aim to carry out a wider screen to test the completeness of expressed human snoRNA annotations. To do so, we have employed StringTie on TGIRT-seq datasets from diverse normal human tissues, cell lines and stem cells to identify expressed genes missing from current annotations, having an intronic or intergenic location with a size between 50 and 200 nucleotides. These data were integrated with immunoprecipitation sequencing studies of core snoRNA binding proteins (PAR-CLIP and eCLIP) as well as RNA-seq studies following the depletion of these same proteins to identify the most likely snoRNA candidates. This methodology allowed us to identify 19 and 5 unannotated genes, which are proposed respectively as box C/D and box H/ACA snoRNAs. Most of these genes have a low conservation level, an intergenic localization, and some have a tissue-specific expression. These results further demonstrate that the annotation of snoRNAs in humans is far from exhaustive, hence the interest of implementing more efficient and more reliable pipelines for their identification.

Replication of the virulence gene ToxB in the fungal plant pathogen Pyrenophora tritici-repentis

Ryan Gourlie[1,2], Megan McDonald[3], Mohamed Hafez[1], and Reem Aboukhaddour[1]

[1]Agriculture and Agri-food Canada, Lethbridge, Alberta, Canada
[2]University of Lethbridge, Lethbridge, Alberta, Canada
[3]University of Birmingham, Edgbaston, Birmingham, UK

Presenting author: ryan.gourlie@agr.gc.ca

plant pathology, gene duplication, transposons, helitron, virulence evolution

Copy-number variation is a major driver of evolution, correlating with increased virulence and adaptability in fungal plant pathogens. In the wheat pathogen *Pyrenophora tritici-repentis* (Ptr), the *ToxB* gene, has varying copy numbers ranging from 0 to 10 copies. We utilized >20 long-read (PacBio RS II) assemblies (Hi-CANU) to understand the replication mechanism of *ToxB* within the Ptr genome. Results revealed that in multi-copy isolates, *ToxB*, along with variable segments of surrounding sequences, exists as tandem unidirectional copies. Distinctive features strongly support the involvement of a Helitron, a class of transposable elements, in *ToxB* replication. Additionally, our analysis showed that *ToxB* resides within a repeat-dense region, rich with transposon activity, including evidence for two different Copia-like transposons disrupting and inactivating the *ToxB* reading-frame (i.e. *toxb*). The region containing *ToxB* is completely absent in isolates lacking the *ToxB* gene. The size of the semi-conserved region may support the presence of a supernumerary chromosome arm, or perhaps an ancient, now defunct, large mobile element. Large regions such as this be created from unequal crossing-over and there is some evidence to suggest that this may have played a role in the duplication of *ToxB* as well.