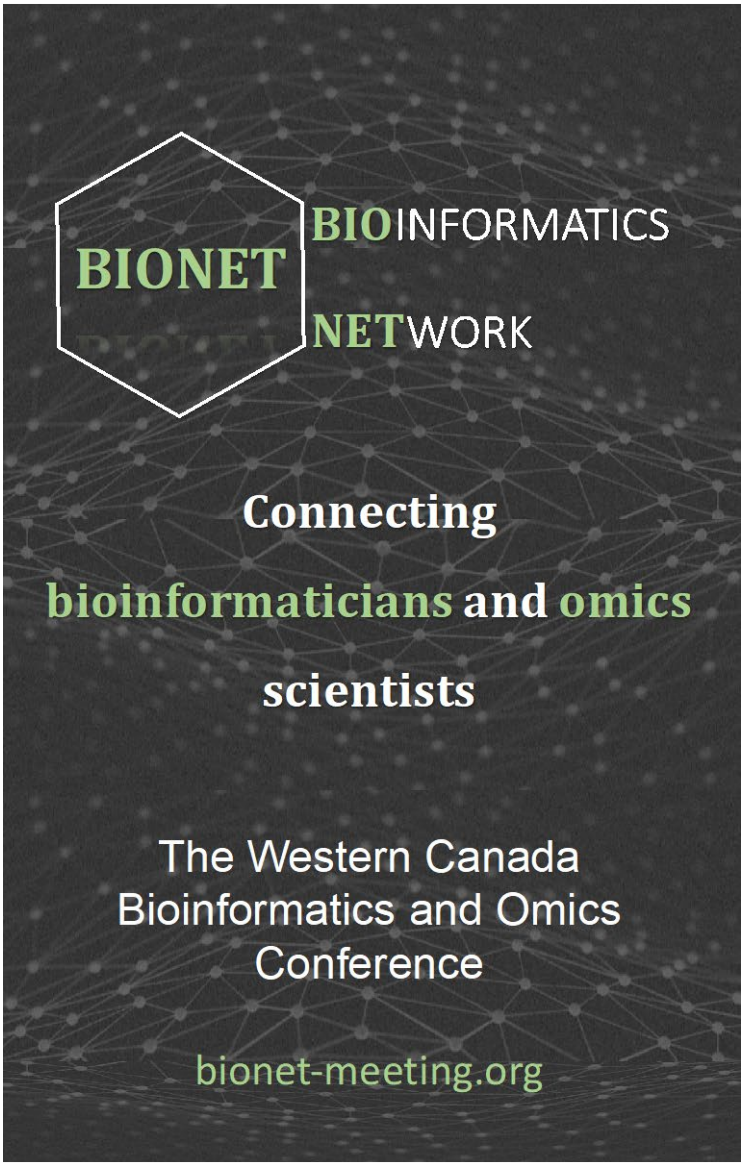
BioNet 2025
Conference Program

May 7th through 9th, the Arthur J. E. Child Comprehensive Cancer Centre, Calgary, Alberta

BIOINFORMATICS NETWORK ALBERTA

BIOINFORMATICS NETWORK PRAIRIE

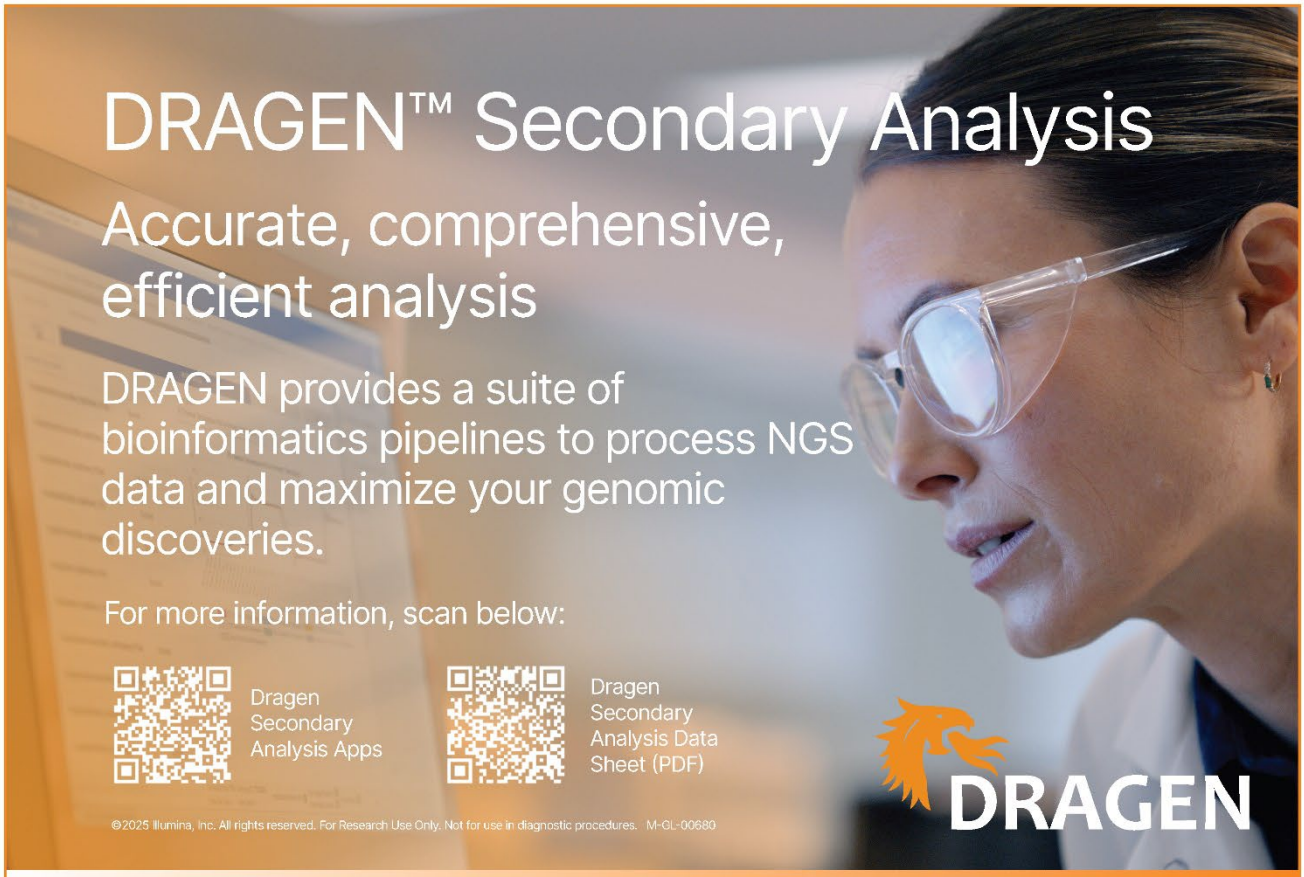


BIOINFORMATICS NETWORK ALBERTA

BIOINFORMATICS NETWORK PRAIRIE

BioNet would like to thank our generous conference partners and sponsors for their support for BioNet 2025:

The University of Calgary, located in the heart of Southern Alberta, both acknowledges and pays tribute to the traditional territories of the peoples of Treaty 7, which include the Blackfoot Confederacy (comprised of the Siksika, the Piikani, and the Kainai First Nations), the Tsuut'ina First Nation, and the Stoney Nakoda (including Chiniki, Bearspaw, and Goodstoney First Nations). The City of Calgary is also home to the Métis Nation of Alberta (Districts 5 and 6).



illumina®

We would like to thank Illumina for their generous support at the **Gold Level Sponsor of BioNet 2025.** This support allows us to keep registration costs low, bring speakers in from across Canada and provide you with the best experience possible.

BIOINFORMATICS NETWORK ALBERTA

BIOINFORMATICS NETWORK PRAIRIE

Welcome Messages

Dr. Athan Zovoilis

Academic Lead, BioNet (on behalf of BIONET 2025 Scientific Organizing Committee)

Welcome to the 5th Annual BioNet Conference—BIONET 2025!

What began as an ambitious initiative in Alberta more than 5 years ago has since evolved into a vibrant, multi-provincial community of bioinformaticians, omics scientists, and data-driven researchers spanning the Prairie provinces and beyond. Our journey from a provincial network to a truly regional hub of computational biology has been marked by shared vision, grassroots collaboration, and a deep commitment to scientific excellence.

Since its launch, BioNet has grown from its inaugural conference of just over 100 participants to a comprehensive ecosystem that now includes a successful annual meeting, training programs, seminar series, placement opportunities, and white papers that have helped articulate the strategic needs of the bioinformatics sector in Western Canada. Along the way, we've cultivated partnerships with leading academic institutions, research hospitals, government agencies, and emerging innovators in agriculture, health, and AI.

Looking ahead, BioNet is actively preparing for a new phase—one marked by national collaboration. We are proud to announce our intention for the development of a long-term partnership with the Canadian Bioinformatics Hub to contribute to a national-level omics and bioinformatics conference. In this model, BioNet will hold its Prairie-focused meetings every second year, strategically alternating with the national event. We are committed to contributing wholeheartedly to both, ensuring continuity, diversity, and synergy within Canada’s growing bioinformatics landscape.

Thank you for being part of this exciting milestone in BioNet’s evolution. Together, we continue to build not only a network—but most importantly a community.

Dr. Pinaki Bose
U of C host (on behalf of BIONET 2025 organizers)

Dear BioNet 2025 Participants,

It is my distinct pleasure to welcome you all to the BioNet 2025 conference, hosted this year at the Arthur Child Comprehensive Cancer Centre at the University of Calgary. We are thrilled to serve as your hosts, and it is truly exciting to see our campus become a vibrant hub for bioinformatics researchers and trainees from across the prairies and beyond.

The BioNet conference represents an opportunity for the local scientific community to foster meaningful collaborations, share innovative research, and build lasting connections. We are eager to strengthen ties and create synergies with scientists and researchers attending from diverse fields and geographic regions.

Bioinformatics plays a pivotal role in the prosperity of our local economy, driving transformative advancements not only in medicine, but also across critical sectors such as agriculture, environmental management, and biotechnology. The innovative work shared at this conference underscores the immense potential bioinformatics has to improve lives and enhance economic growth in Alberta and beyond.

Thank you all for joining us. I look forward to a stimulating and fruitful conference filled with outstanding science, passionate discussions and new partnerships.

Warm regards,

Dr. Pinaki Bose
Local Host, BioNet Alberta Conference
Associate Professor
Department of Oncology
Arthur Child Comprehensive Cancer Centre
University of Calgary

BioNet 2025 Organizers

Scientific Committee

Dr. Pinaki Bose, University of Calgary
Dr. Athan Zovoilis, University of Manitoba
Dr. Elisabeth Richardson, Mount Royal University
Dr. Angeliki Pantazi, University of Manitoba
Dr. Emily Herman, University of Alberta
Dr. Srijak Bhatnagar, Athabasca University
Dr. Jason de Koning, University of Calgary
Dr. Paul Gordon, University of Calgary
Dr. Eric Merzetti, University of Manitoba / Genome Alberta

Event Committee

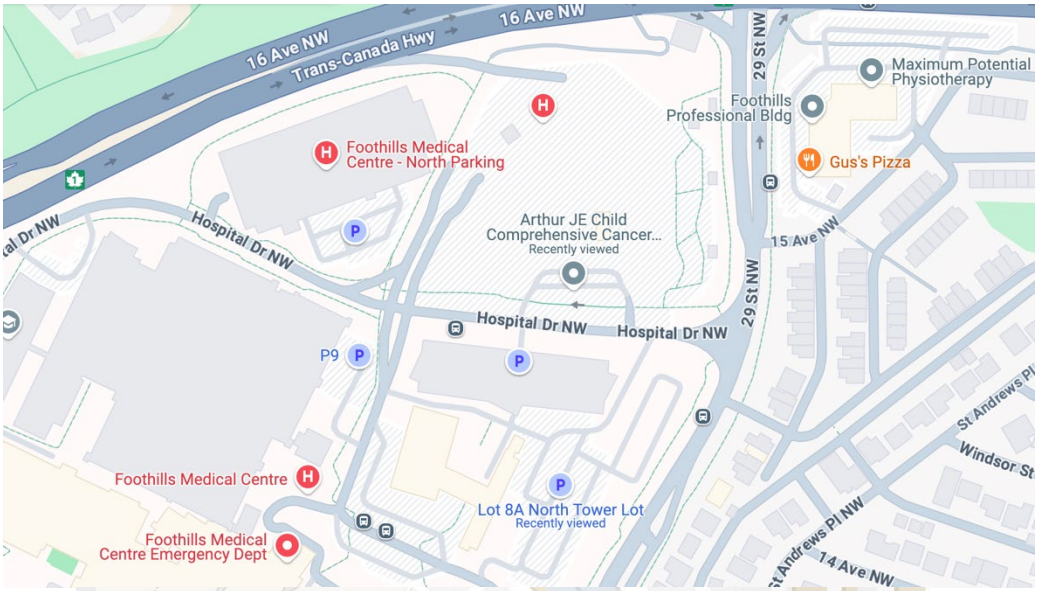
Dr. Pinaki Bose, University of Calgary
Dr. Athan Zovoilis, University of Manitoba
Dr. Eric Merzetti, University of Manitoba / Genome Alberta
Dr. Sani e-Zehra Zaidi, University of Manitoba
K M Tahsin Hassan Rahit, University of Calgary
Dr. Ayan Chanda, University of Calgary
Riya Roy, University of Manitoba
Alex Sizykh, University of Manitoba
Parisa Rayomand, University of Manitoba
Alyona Gerasimova, University of Manitoba
Liam Mitchell, University of Manitoba

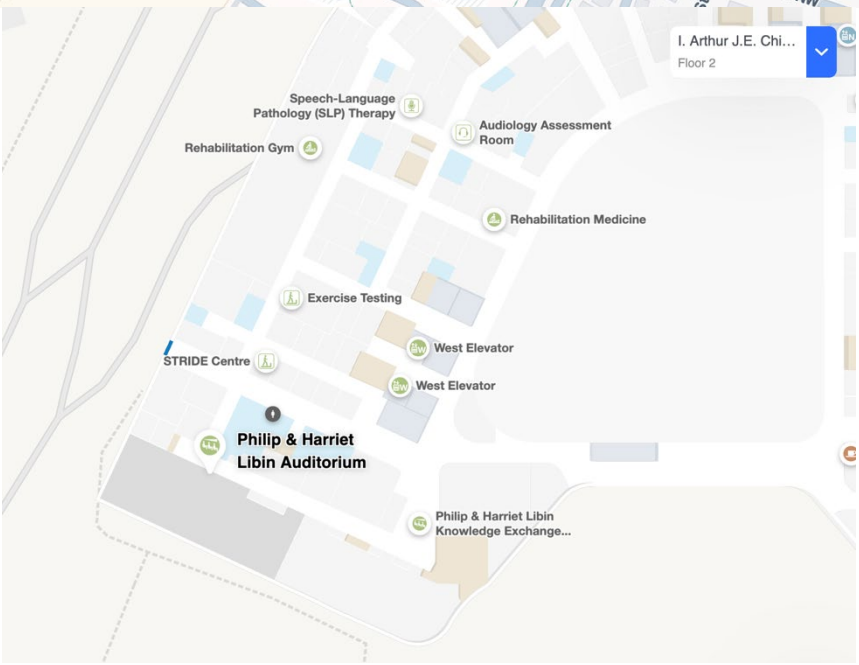
Poster Judges

Dr. Elisabeth Richardson, Mount Royal University
Dr. Sara Smith, Mount Royal University
Dr. Emily Herman, University of Alberta
Dr. Srijak Bhatnagar, Athabasca University
Dr. Mark Berjanskii, University of Alberta
Dr. Gabrielle Scheffer, Genome Alberta

Venue

Philip & Harriet Libin Knowledge Exchange Centre and Auditorium,
Arthur JE Child Comprehensive Cancer Centre.
3395 Hospital Drive NW,
Calgary, Alberta, T2N 5G2
1-844-465-6330 (587-231-3100)





BioNet 2025 At A Glance

Wednesday, May 7, 2025 1:00PM-7:00PM	**The Philip & Harriet Libin Knowledge Exchange Centre and Auditorium** 2nd Floor, Arthur J Child Comprehensive Cancer Centre
1:00pm-2:00pm	**Registration / Doors Open**
2:00pm-2:30pm	Welcome and Opening Remarks
2:30pm-3:45pm	**Session 1: Disease Omics**
3:45pm-4:15pm	Refreshment Break
4:15pm-5:30pm	**Session 2: Health Omics**
5:30pm-6:00pm	**Opening Plenary: Carina Butterworth**
6:00pm-7:00pm	Dinner

Thursday, May 8, 2025 8:00AM-6:00PM	**The Philip & Harriet Libin Knowledge Exchange Centre and Auditorium** 2nd Floor, Arthur J Child Comprehensive Cancer Centre
8:00am-9:00am	Breakfast
9:00am-9:25sm	**Plenary: Dr. Michelle Brazas**
9:30am-10:50am	**Session 3: Cancer Omics**
10:50am-11:10am	Refreshment Break
11:10am-12:10pm	**Session 4: Panel Discussion**

12:15pm-12:30pm	**Sponsor Talk: Illumina**
12:30pm-1:15pm	Lunch
1:15pm-2:45pm	**Session 5: Poster Presenter Flash Talk**
2:45pm-3:00pm	Refreshment Break
3:00pm-4:00pm	Poster presentation (Even numbered)
4:00pm-5:00pm	Poster presentation (Odd numbered)
5:00pm-6:00pm	**Keynote: Dr. Ali Bashashati**
7:00pm-10:00pm	**The Bionet Networking and Work placement Program Mixer Event**

Friday, May 9, 2025
8:00AM-1:00PM

The Philip & Harriet Libin Knowledge Exchange Centre and Auditorium
2nd Floor, Arthur J Child Comprehensive Cancer Centre

8:00am-9:00am	Breakfast
9:00am-9:30am	**Co-Talk: Drs Elisabeth Richardson & Sara Smith**
9:30am-10:30am	**Session 6: Environmental Omics**
10:35am-11:05am	Refreshment Break
11:05am-12:20pm	**Session 7: Bioinformatics in Livestock and Agriculture**
12:30pm-1:00pm	Awards / Closing Remarks

BioNet 2025 – Full Program

Wednesday, May 7, 2025

1:00PM-7:00PM

Check in and badge pickup starting at 1:00pm in the Atrium of the Philip & Harriet Libin Knowledge Exchange Centre and Auditorium on floor 2 of the Arthur J Child Comprehensive Cancer Centre

Opening Ceremony

2:00 – 2:15pm	**Dr. Athan Zovoilis** Academic Lead, BioNet	
2:15 – 2:30pm	**Dr. Eric Merzetti** Network Manager, BioNet, Co-Founder BioNet Prairie, Prairies Coordinator Canadian Bioinformatics Hub, CCMB Core	

Session 1: Disease Omics

Chair: Dr. Ryan Mercer, Alberta Innovates

2:30 – 2:45pm	**Dr. James Wasmuth, University of Calgary** "Molecular Doppelgängers: Detecting mimicry in host-pathogen interactions."	**4**
2:50 – 3:05pm	**Dr. Kolapo Oyebola, Centre for Genomic Research in Biomedicine, Nigeria** "Machine Learning Insights on the Effectiveness of Non-Pharmaceutical Interventions against COVID-19 in Nigeria"	**6**
3:10 – 3:25pm	**Dr. Galen Wright, University of Manitoba** "Unbiased human genomic characterization of polyglutamine disorder genes to inform therapeutic strategies"	**7**
3:30 – 3:45pm	**Dr. Ayooluwa J Boalji, National Microbiology Laboratory, Winnipeg** "In silico analyses identify sequence contamination thresholds for Nanopore-generated SARS-CoV-2 sequences"	**8**
3:45 – 4:15pm	**Refreshment Break – PHLA Atrium**	

Session 2: Omics in Health

Chair: Dr. Georgia Balsevich, Genome Alberta

4:15 – 4:30pm	**Dr. Mohamed Helmy, Vaccine and Infectious Disease Organization, University of Saskatchewan** "wiseFlu: A System for Weekly Surveillance of Influenza A (H5N1) Sequence Evolution to Support Human Health through Early Detection and Vaccine Candidate Identification."	**9**
4:35 – 4:50pm	**Dr. Igor Kovalchuk, University of Lethbridge** "Analysis of sex-specific differences in DNA methylation patterns associated with aging"	**10**
4:55 – 5:10pm	**Dr. Quan Long, University of Calgary, University of Calgary** "Tissue-specific transfer learning redirects comprehensive models to target disorders"	**11**
5:15 – 5:30pm	**Dr. Diogo Pellegrina, Vaccine and Infectious Disease Organization, University of Saskatchewan** "Pathway De-nesting: A novel method for correcting pathway enrichment analyses by taking ontology structures into account"	

Opening Plenary

Chair: Dr. Athan Zovoilis, University of Manitoba

5:30 – 6:00pm	**Carina Butterworth, Southern Alberta Institute of Technology** "Don't Become a Headline: Ethical and Inclusive AI"	**3**
6:00 – 7:00pm	**Supper**	

Thursday, May 8, 2025

8:00AM-6:00PM

8:00 – 9:00 am **Breakfast**

Plenary Talk

Chair: Dr. Athan Zovoilis, University of Manitoba

8:45 – 9:10 am	**Dr. Michelle Brazas, Ontario Institute for Cancer Research** "Bridging the Bioinformatics and Computational Biology Training Gap"	**2**

Session 3: Cancer Omics

Chair: Dr. Tom Finn, Genome Alberta

9:15 – 9:30am	**Dr. Michael Charette, Brandon University** TBD	
9:30 – 9:45am	**Dr. Pinaki Bose, University of Calgary** "Spatial Profiles of Primary and Recurrent Head and Neck Cancers"	**13**
9:50 – 10:05am	**Dr. Nehal Thakor, University of Lethbridge** "Establishing eukaryotic initiation factor 5B (eIF5B) as a prognostic biomarker and therapeutic target for oral squamous cell carcinoma"	**14**
10:10 – 10:25am	**Dr. Britt I. Drögemöller, University of Manitoba** "Leveraging large-scale datasets and single cell omics data to develop a polygenic score for cisplatin-induced ototoxicity"	**15**
10:30 – 10:40am	**Dr. Ayan Chanda, University of Calgary** "Multi-omic Dissection of Lymph Node Metastasis in Oral Squamous Cell Carcinoma"	**16**
10:40 – 11:05am	**Refreshment Break**	

Session 4: Panel Discussion

Applications of AI in bioinformatics and the rapidly changing landscape
11:05-12:10am
Chair: Dr. Athan Zovoilis, University of Manitoba

Speakers:
Dr. Michelle Brazas, Ontario Institute for Cancer Research
Carina Butterworth, Southern Alberta Institute of Technology
Dr. Steven Jones, BC Cancer Research Institute
Dr. Raja B Singh, BioAro

Sponsor Talk: Illumina

Chair: Dr. Eric Merzetti, BioNet / Genome Alberta

12:15 – 12:30pm	**Sidki Bouslama, Illumina** "Advancements in Illumina Informatics"
12:30 – 1:15pm	**Lunch**

Session 5: Flash Talks

Please welcome our poster presenters who will be given 3 minutes and the option of presenting 1-slide each to the audience in advance of the formal poster session.

1:15-2:45pm

Chair: Dr. Emily Herman, University of Alberta

Presenter	Affiliation	Poster #
Daniel Armando Salazar Aleman	University of Calgary	3
Amy Banks	University of Calgary	4
Charlotte Bourbon	University of Calgary	6
Yuxuan Chen	University of Calgary	7
David Enoma	University of Calgary	9
Ariel Ghislain Kemogne Kamdoum	University of Calgary	10
Mawra Gohar	University of Calgary	11
Kevin Joannou	University of Alberta	13
Emily Kingdon	Mount Royal University / University of Alberta	14
Dr. Yi Lan	University of Calgary	16
Ananda Mahafujul	University of Saskatchewan	20
Dr. Gen Morinaga	University of Calgary	23
Keyhan Najafian	University of Saskatchewan	24
Sweta Rai	University of Lethbridge	25
Tayab Soomro	University of Saskatchewan	28
Jonas Stadfeld	University of Calgary	29
Malcolm Todd	University of Saskatchewan	31
Kiana Vadiat	Mount Royal University	32
Dinghao Wang	University of Calgary	33

2:45 – 3:00pm **Refreshment Break**

The BioNet 2025 Poster Session – Atrium

3:00 – 4:00pm	Even Numbered Posters Present
4:00 – 5:00pm	Odd Numbered Posters Present

Keynote Talk – Dr. Ali Bashashati, University of British Columbia

Chair: Dr. Athan Zovoilis, University of Manitoba

5:00-6:00pm Machine Learning in Cancer Biomarker Discovery: Promise, Prospects, and Progress

7:00 – 10:00pm **The BioNet Networking and Work Placement Program Mixer Event**

Join us for the BioNet 2025 Industry Mixer and Partnership event on Thursday, May 8, from 7-10 pm. An ideal opportunity for local companies to connect with up-and-coming talent and for early-career professionals to network with industry veterans. Industry representatives can discover new expertise and workflows, while trainees and recent graduates can identify job opportunities in rapidly growing fields.

- Networking with industry professionals and academic experts in bioinformatics, genomics, proteomics, and related fields
- Career opportunities for recent graduates and trainees eager to enter the field
- Collaboration discussions between industry partners and bioinformatics specialists looking to tackle tomorrow's scientific challenges
- Insightful conversations on industry trends, the skills in demand, and how bioinformatics is shaping the future of healthcare, agriculture, and beyond

Whether you're looking for talent or new career opportunities, this event provides a unique space for forging professional connections, building partnerships, and shaping the future of bioinformatics.

Friday, May 9, 2025

8:00AM-1:00PM

Chair: Dr. Eric Merzetti

9:00 – 9:30am	**Drs. Elisabeth Richardson and Sara Smith, Mount Royal University** "Bioinformatics as an Essential Component of a Biology Undergraduate Degree"	

Session 6: Environmental Omics

Chair: Dr. Elisabeth Richardson

9:35 – 9:50am	**Dr. Marc Strous, University of Calgary** "The role of groundwater microbiota in the natural bioremediation of nitrate."	**19**
9:55 – 10:10am	**Dr. Huiqing Yeo, University of Calgary** "The origin and spread of Culex pipiens in Alberta"	**20**
10:15 – 10:30 am	**Dr. Tulika Bhardwaj, University of Calgary** "Optimizing viral detection in wastewater metatranscriptomes through targeted depletion of Tobamoviruses and ribosomal RNAs"	**21**
10:35 – 11:05 pm	**Refreshment Break**	

Session 7: Bioinformatics in Livestock and Agriculture

Chair: Rodrigo Ortega-Polo, Agriculture and Agri-Food Canada

11:05 – 11:20am	**Dr. Nathaniel Lim, Agriculture and Agri-Food Canada** "Leveraging artificial intelligence and the cloud in advancing agricultural crop research"	**23**
11:25 – 11:40am	**Dr. Emily Herman, University of Alberta** "Beyond SNPs: Developing a structural variant detection workflow and associated database for Holstein cattle"	**25**
11:45 – 12:00pm	**Dr. Dongyan Niu, University of Calgary** "Decoding poultry threats: unveiling sequencing types, serogroups, and antimicrobial resistance of avian pathogenic *e. coli* in Alberta, Canada"	**26**
12:05 – 12:20pm	**Dr. Lael Barlow, University of Alberta** "A genomic admixture testing method for bison conservation"	**28**
12:30 – 1:00 pm	**Closing Ceremonies – Awards / Final Comments**	

BioNet 2025 – Poster Presentations

BioNet 2025 Invited Speakers

Dr. Ali Bashashati, University of British Columbia

Keynote Talk

Machine Learning in Cancer Biomarker Discovery: Promise, Prospects, and Progress

Dr. Bashashati is an Associate Professor at UBC, Principal Investigator at VCHRI, and Director of AI and Bioinformatics in BC's Ovarian Cancer Research Program.

Dr. Bashashati's research area lies at the interface between computational, engineering and biomedical sciences. He is interested in developing machine learning, statistical and signal processing algorithms and software infrastructure to combine various sources of omics and imaging data with major emphasis on discovering novel complex biological information related to different diseases. His research is specifically focused on ovarian and breast cancers as well as lymphoid malignancies and how these cancers evolve and respond to therapies. He has published extensively in cancer genomics, bioinformatics, computational biology and brain computer interface fields and his papers have appeared in top-tier journals such as Nature and Nature Genetics.

Dr. Michelle Brazas, Ontario Institute for Cancer Research

Plenary Talk

Bridging the Bioinformatics and Computational Biology Training Gap

Dr. Brazas is the Director of Adaptive Oncology at OICR and Scientific Director of Bioinformatics.ca, where she promotes bioinformatics training and community building. She co-founded the Toronto Bioinformatics User Group (TorBUG) and plays a key role in advancing bioinformatics education as an executive with the ISCB and a member of GOBLET, a global organization supporting bioinformatics learning and training.

Carina Butterworth, P Eng, Southern Alberta Institute of Technology

Plenary Talk

Don't Become a Headline: Ethical and Inclusive AI

Carina holds a BSc, MSc in Geomatics Engineering, and a Certificate for Emerging Leaders from the University of Calgary. Currently pursuing a Ph.D. in Biomedical Engineering, her research focuses on improving breast cancer imaging using microwave technology. An award-winning instructor in SAIT's Geomatics Engineering Technology program, she was the recipient of the 2021 Saitsa Teaching Excellence Award, National Geomatics Leader in Innovation Award, the SAIT Brand Original Award (2016) and the International Women's Day Catalyst Award (2020).

BioNet 2025
Conference Speakers

Molecular Doppelgängers: Detecting mimicry in host-pathogen interactions

James Wasmuth[1], Kaylee Rich[1], Shruti Srivastava[1], Viraj Muthye[1]

[1] Faculty of Veterinary Medicine, University of Calgary, Calgary, Canada

Presenting author: jwasmuth@ucalgary.ca

Keywords: Molecular mimicry, Host-pathogen interactions, Sequence alignment, Structural alignment, Parameter optimisation, Tuberculosis, Malaria

Pathogens have evolved sophisticated strategies to subvert host defenses and promote infection. One strategy is molecular mimicry: the structural or functional imitation of host biomolecules by the pathogen. These pathogen-derived molecules may disrupt immune detection and manipulate host signaling. However, identifying mimicry events remains challenging, particularly when similarities are subtle or localized to short protein regions.

To address this, we developed *mimicDetector*, a computational tool that detects molecular mimicry through sequence and structural analyses [1]. Building on existing sequence similarity detection methods, our pipeline integrates optimized short-sequence alignment (using the PAM30 substitution matrix for sensitive short-motif matching) and bitscore-based filtering to prioritize biochemically meaningful mimics. Applied to 32 pathogens of global health importance—including parasites, bacteria, and fungi—*mimicDetector* identified novel mimics such as helminth proteins resembling human complement regulators and a *Leishmania infantum* protein mimicking Reticulon-4, which modulates immune cell trafficking.

We applied *mimicDetector* to *Mycobacterium tuberculosis* (*Mtb*), the causative agent of tuberculosis, revealed potentially novel *Mtb*-human protein interactions. By comparing virulent and attenuated Mtb strains and other, non-pathogenic, *Mycobacterium* species, we identified 122 human proteins as potential targets. These targets included proteins involved in the regulation of the phagolysosome, the host immune cell in which *Mtb* makes its home.

Expanding beyond sequence, we have taken exploited advances in structural bioinformatics, including AlphaFold and Foldseek [2]. In *Plasmodium falciparum*, the causative agent of human malaria, this approach revealed 44 structural mimics of human proteins, three of which had known roles in immune evasion. Notably, cytoskeletal and extracellular matrix proteins were frequent targets, while immune-related proteins were underrepresented, a possible adaptation to avoid immune recognition.

By combining sequence and structural approaches, *mimicDetector* provides a scalable framework to dissect host-pathogen interactions. These findings not only advance our understanding of molecular mimicry but also support future development of therapeutic strategies or vaccines targeting critical pathogen mimicry nodes.

REFERENCES

1. https://doi.org/10.7717/peerj.16339
2. https://doi.org/10.3389/fpara.2023.1162697

Machine Learning Insights on the Effectiveness of Non-Pharmaceutical Interventions against COVID-19 in Nigeria

*Kolapo Oyebola [1,2], Funmilayo Ligali [1,2], Afolabi Owoloye [1,2], Adesola Musa[2], Oluwagbemiga Aina[2] and Babatunde Salako[2]

[1] Centre for Genomic Research in Biomedicine, Mountain Top University, Ibafo, Nigeria
[2] Nigerian Institute of Medical Research, Lagos, Nigeria

Presenting author: oyebolakolapo@yahoo.com

Keywords: Clustering Algorithms, COVID-19, Epidemic Responses, Non-pharmaceutical Interventions, Outbreaks, Regression Analysis

Background: The lack of effective pharmacological measures during the early phase of COVID-19 pandemic prompted the implementation of non-pharmaceutical interventions (NPIs) as initial mitigation strategies. The impact of these NPIs on COVID-19 in Nigeria is not well-documented. This study sought to assess the effectiveness of NPIs to support future epidemic responses.
Methodology: Daily COVID-19 cases and deaths were analysed using smoothed variables to identify transmission trends. Regression analysis and clustering algorithms were applied to evaluate the impact of each NPI.
Results: Multiple transmission peaks were reported, with the highest smoothed daily new cases (~1,790) observed around 29 December 2021 and smoothed daily new deaths (~23) peaking around 8 September 2021. NPIs such as public transport (coefficient value = -166.56, P = 0.01) and workplace closures (coefficient value = -150.06, p = 0.01) strongly correlated with reduced cases. This finding highlights the importance of mobility control and non-essential workplace management in slowing infection transmission during an outbreak. Public transport restrictions (coefficient value = -2.43, $p < 0.001$) also had a direct effect on death reduction.
Conclusions: Public transport restrictions and workplace closures correlated with reductions in cases and deaths. These findings can guide future pandemic responses to enhance favourable public health outcomes.

Unbiased human genomic characterization of polyglutamine disorder genes to inform therapeutic strategies

Kevin Lucy Namuli,[1,2] Britt I. Drögemöller,[3] Galen E.B. Wright[1,2,3]

[1]Department of Pharmacology and Therapeutics, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada.

[2]PrairieNeuro Research Centre, Kleysen Institute for Advanced Medicine, Health Sciences Centre and Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada

[3]Department of Biochemistry and Medical Genetics, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Canada.

Presenting author: galen.wright@umanitoba.ca

Keywords: polyglutamine disorders, repeat expansion disorders, genetic modifiers, human genetics, drug targets, genome-wide association studies

Background: Polyglutamine (polyQ) disorders, such as Huntington disease (HD) and several spinocerebellar ataxias, are severe neurological disorders caused by repeat expansions of the glutamine codon. These conditions lack effective treatments, with therapeutic research focused on pathogenic gene knockdown.

Objectives: We aimed to profile these genes using diverse human genomic data to guide therapeutic strategies by identifying new biology and assessing off-target effects of knocking these genes down.

Methods: We conducted an unbiased phenome-wide study to identify human traits and diseases linked to polyQ disorder genes (Open Targets L2G>0.5). Network analyses explored shared trait associations and overlapping biological processes among these genes. Lastly, we assessed the theoretical druggability of polyQ disorder genes using recently identified features predictive of clinical trial success and compared them to repeat expansion (HD) modifier genes.

Results: We identified 215 human phenotype/polyQ disorder gene associations from 3,095 studies, indicating potential adverse effects from gene knockdown. Shared trait associations among polyQ disorder genes suggested overlapping biological processes despite distinct functions. Drug target profile analysis revealed unfavorable risk profiles for polyQ disorder genes, particularly *ATN1*, *ATXN1*, *ATXN7*, and *HTT*, due to genomic features such as constraint, molecular interactions, and tissue specificity. PolyQ disorder genes also showed significantly more safety-related risks than HD genetic modifier genes (P=7.03x10^{-3}).

Conclusion: Our analyses emphasize the pleiotropic nature of polyQ disorder genes, highlighting their potential risks as drug targets due to safety concerns. These findings reinforce the importance of exploring alternate therapeutic strategies, such as targeting genetic modifier genes, to mitigate these challenges.

In silico analyses identify sequence contamination thresholds for Nanopore-generated SARS-CoV-2 sequences

Ayooluwa J. Boalji[1,2,3], Ana T. Duggan[1]

[1]Public Health Agency of Canada, National Microbiology Laboratory, Winnipeg, Canada
[2] Cadham Provincial Laboratory, Winnipeg, Canada
[3]University of Manitoba, Department of Biochemistry and Medical Genetics, Winnipeg, Canada

Presenting author: ayooluwa.bolaji@phac-aspc.gc.ca

Keywords: Genomics, NGS, ONT, in silico analyses, SARS-CoV-2, phylogenetics, SNVs, lineage calls, genome coverage, genome completeness

The SARS-CoV-2 pandemic has brought molecular biology and genomic sequencing into the public consciousness and lexicon. With an emphasis on rapid turnaround, genomic data informed both diagnostic and surveillance decisions for the current pandemic at a previously unheard-of scale. The surge in the submission of genomic data to publicly available databases proved essential as comparing different genome sequences offers a wealth of knowledge, including phylogenetic links, modes of transmission, rates of evolution, and the impact of mutations on infection and disease severity. However, the scale of the pandemic has meant that sequencing runs are rarely repeated due to limited sample material and/or the availability of sequencing resources, resulting in the upload of some imperfect runs to public repositories. As a result, it is crucial to investigate the data obtained from these imperfect runs to determine whether the results are reliable prior to depositing them in a public database. Numerous studies have identified a variety of sources of contamination in public next-generation sequencing (NGS) data as the number of NGS studies increases along with the diversity of sequencing technologies and procedures. For this study, we conducted an in silico experiment with known SARS-CoV-2 sequences produced from Oxford Nanopore Technologies sequencing to investigate the effect of contamination on lineage calls and single nucleotide variants (SNVs). A contamination threshold below which runs are expected to generate accurate lineage calls and maintain genome-relatedness and integrity was identified. Together, these findings provide a benchmark below which imperfect runs may be considered robust for reporting results to both stakeholders and public repositories and reduce the need for repeat or wasted runs.

***wiseFlu*: A System for Weekly Surveillance of Influenza A (H5N1) Sequence Evolution to Support Human Health through Early Detection and Vaccine Candidate Identification**

Mahafujul Ananda[1], Matthew Rogers[2], Lingling Jin[1], Mohamed Helmy[1~4]

[1] Department of Computer Science, University of Saskatchewan, Saskatoon, SK, Canada
[2] Vaccine and Infectious Diseases Organization (VIDO), University of Saskatchewan, Saskatoon, SK, Canada
[3] School of Public Health, University of Saskatchewan, Saskatoon, SK, Canada
[4] Department of Computer Science, Lakehead University, Thunder Bay, ON, Canada

Presenting author: mohamed.helmy@usask.ca

Keywords: Avian Flu, Genomic Surveillance, Influenzas A, Bioinformatics, Systems Biology, AI, Phylogenetics, Vaccine Development, Public Health

Avian influenza A (H5N1) is a growing threat to human health due to its rapid mutation rate and zoonotic transmission potential. Recent human infections in Canada and the US, including one fatality, highlight the virus's capacity to spill over from livestock into humans. Saskatchewan faces a heightened risk due to its major role in Canada's poultry and cattle industries and its location along key migratory bird pathways, making it vulnerable to viral introduction and spread. Despite the risks, current surveillance systems, such as FluWatch and FluWatchers, lack real-time genomic analysis capabilities necessary for detecting emerging strains and informing vaccine design.

To address this critical gap, we developed ***wiseFlu***, a platform for the weekly surveillance of H5N1 sequence evolution. The system automatically collects new viral genomic sequences from platforms like GISAID, applies quality control measures, and performs phylogenetic analyses to detect concerning mutations. ***wiseFlu*** identifies variants with potential for increased transmissibility, immune evasion, or virulence based on reference data from the past decade.
WiseFlu's web portal provides real-time access to processed data and interactive visualization tools, allowing virologists, public health officials, and policymakers to track viral evolution and make informed decisions. The system generates automated weekly reports highlighting newly identified high-risk strains and potential vaccine escape variants. These reports directly support vaccine development efforts at the Vaccine and Infectious Disease Organization (VIDO) by identifying candidate strains for experimental validation.

By offering near real-time genomic surveillance, ***wiseFlu*** strengthens our capacity to detect and respond to H5N1 outbreaks early, mitigating risks to human health, agriculture, and food security. The platform's scalable framework is designed for rapid adaptation to other influenza subtypes and emerging pathogens, making it a vital tool for pandemic preparedness. The wiseFlu system is available at (https://wiseflu.usask.ca/).

Analysis of sex-specific differences in DNA methylation patterns associated with aging

Bo Wang[1,2], DongPing Li[1,2], Yaroslav Ilnytskyy[1,2], Olga Kovalchuk[1,2] and Igor Kovalchuk[1,2]

[1] Department of Biological Sciences, University of Lethbridge, Lethbridge, Alberta, T1K 3M4, Canada;
[2] Pathway Rx., 2 Fortress Rise SW, Alberta, T3H 4Z2, Canada

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

Keywords: aging, DNA methylation, biological clock, sex-specific aging-associated changes, methylation-sensitive PCR

Aging is a gradual deterioration of molecular and cellular processes, resulting in the decline of multiple organ systems culminating in frailty and death. Chronic diseases are often associated with and aggravated by aging. Does aging cause the organism to be predisposed to chronic conditions, or do chronic conditions trigger accelerated aging, or both? Also, do males and females age in the same manner? Are age-related symptoms of various disease different between sexes? Answers to these questions can be obtained by analyzing omics data from large cohorts.

Analysis of DNA methylation is one such omics analysis. DNA methylation changes in response to lifestyle and environment, allowing to capture a snapshot of our health. Analysis of multiple data points in a longitudinal cohort allows analysis of DNA methylation changes over time and thus allows to correlate those changes with health/disease trajectory. DNA methylation also correlates well with human age. Some humans age prematurely, and their biological age appears older than their chronological age. In contrast, some age slower, reflecting either genetic predisposition to rich older age or health lifestyle, or both. Finally, knowing how the health of an individual changes with time should allow us to predict health changes based on where on a healthy aging trajectory this individual is positioned.

In this work, we have profiled large sets cohorts with DNA methylation and chronological age data. We various bioinformatics approaches to identify sex-specific patterns of changes in DNA methylation correlating with aging. While most of the genes/pathways overlapped, we found some unique significant differences in the way males and females age. We then selected top five gene candidates with highest degree of correlation with age and attempted to establish PCR-based assay for the analysis of biological age. We developed and used TaqMan methylation-sensitive PCR to analyze methylation levels. We are currently validating this assay on huma DNA samples.

Tissue-specific transfer learning redirects comprehensive models to target disorders

Quan Long

University of Calgary, Calgary, Canada

Presenting author: quan.long@ucalgary.ca

Keywords: Transfer learning, Cancer genomics, Transcription factor, ChIP-seq, Transcriptome-wide association study

Machine learning (ML) has proven successful in biological data analysis. However, it may require massive training data. To allow broader use of ML in the full spectrum of biology and medicine, including sample-sparse domains, it is critical to re-direct established models to specific tasks by add-on training via a moderate sample. Transfer learning (TL), a technique migrating pre-trained models to new tasks, fits in this requirement. Here, by TL, we redirected Enformer, a comprehensive model trained by massive data of 5,313 tracks of omics, to breast and prostate cancers using tissue-specific transcription factor (TF) ChIP-seq data. Its performance has been validated through statistical accuracy of predictions, annotation of genetic variants, and mapping of genes associated with cancers. By allowing the flexibility of adding dedicated training data, our TL protocol unlocks future discovery within specific domains with moderate add-on samples by standing on the shoulders of giant models.

Pathway De-nesting: A novel method for correcting pathway enrichment analyses by taking ontology structures into account

Diogo Pellegrina[1] and Mohamed Helmy[1,2,3,4,5,6]

[1] Vaccine and Infectious Diseases Organization (VIDO), University of Saskatchewan, Saskatoon, SK S7N 5A2, Canada
[2] Vaccinology and Immunotherapeutics Program, School of Public Health, University of Saskatchewan, Saskatoon, SK S7N 5A2, Canada
[3] Department of Computer Science, University of Saskatchewan, Saskatoon, SK S7N 5A2, Canada
[4] Department of Computer Science, Lakehead University, Thunder Bay, ON P7B 5E1, Canada
[5] Department of Computer Science, Idaho State University, Pocatello, ID 83209, USA
[6] Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), Singapore 138632, Singapore

Presenting author: diogo.pellegrina@usask.ca

Keywords: Bioinformatics, Pathway Enrichment Analysis, Methods, Algorithm, Sampling

Pathway enrichment analysis is a powerful tool that measures how many more genes of interest a previously annotated gene ontology has than expected. It is commonly used by researchers to describe a list of selected genes in easily interpretable terms. In current pathway enrichment methods each pathway is tested independently, without taking into consideration that these pathways are structured in hierarchical terms where the largest terms (composed of thousands of genes) contain smaller terms that partially intersect each other and in succession contain even smaller terms. This Nested pattern continues until the smallest terms that may have less than a dozen genes. Because of that some highly significant pathways will inevitably share all, or some, of their genes with other pathways that would be also considered enriched just based on the genes in its intersection. This hitchhiker affects all terms that share a subset of their genes, and inflates their significance, often giving them better statistical significance than other relevant terms, therefore misleading the interpretation of the results. This effect also produces results that are very repetitive with many pathways that can have more than 90% of their genes in common. We created PathwayDeNester an algorithm that takes the other pathways into consideration and assigns pathways a p-value based on the likelihood that it was enriched by its own and not by the genes in the intersection with more significant pathways. This is achieved by testing each pathway (A) with each of the other pathways with higher significance (B), it takes into account the size of pathway A, its number of selected genes, the number of genes it has in common with pathway B, and the number of selected genes that fall in this intersection. In this algorithm the same number of genes on pathway A is randomly selected, if the amount that would fall in the intersection with pathway B is consistently smaller than what is observed then we can consider that pathway A has its enrichment dependent on pathway B, and therefore should not be considered as a representation of the dataset. We benchmarked PathwayDeNester with simulated and real datasets, filtering out between ~40% and ~90% of the imputed pathways.

Spatial Profiles of Primary and Recurrent Head and Neck Cancers

Pinaki Bose[1,2,3,4]

[1]Department of Oncology, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada.
[2]Department of Biochemistry and Molecular Biology, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada.
[3]Department of Pathology and Laboratory Medicine, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada.
[4]Ohlson Research Initiative, Arnie Charbonneau Cancer Institute, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada.

Corresponding author: Pinaki Bose – pbose@ucalgary.ca

Head and neck squamous cell carcinoma (HNSCC) represents a heterogeneous group of malignancies arising from the epithelial lining of the oral cavity, pharynx, and larynx. Despite advances in multimodal therapies combining surgery, radiation, and chemotherapy, HNSCC is associated with high rates of recurrence, therapeutic resistance, and poor overall survival. Tumour heterogeneity, characterized by distinct molecular subpopulations and spatial variations within the tumour microenvironment (TME), significantly contributes to treatment failure and disease progression. Spatial transcriptomics (ST), which integrates histological imaging with spatially resolved gene expression data, provides an opportunity to characterize the complex molecular architecture of cancer. In this talk, I will describe the results from ST profiling of primary and recurrent HNSCC and specific gene expression patterns in the leading edge and tumour core impact immune modulation.

Establishing eukaryotic initiation factor 5B (eIF5B) as a prognostic biomarker and therapeutic target for oral squamous cell carcinoma

Pavan Lakshmi Narasimha [1], Jinay Patel [1], Ayan Chanda [2], Veda Hegde [1], Hannah Stephenson [1], *Seyed Mehdi Jafarnejad [3], *Pinaki Bose [2, 4], and *Nehal Thakor [1, 4]

[1] *Department of Biological Sciences, University of Lethbridge, AB, Canada.*
[2] *Department of Biochemistry and Molecular Biology, University of Calgary, AB, Canada*
[3] *School of Medicine, Dentistry and Biomedical Sciences, Queen's University, Belfast, UK*
[4] *Arnie Charbonneau Cancer Institute, Cumming School of Medicine, University of Calgary, AB, Canada*

** Co-corresponding authors: sm.jafarnejad@qub.ac.uk; pbose@ucalgary.ca; nthakor@uleth.ca*

Presenting author: nthakor@uleth.ca

Keywords: mRNA translation, non-canonical translation, HNSCC, OSCC, eukaryotic initiation factors (eIFs)

Oral squamous cell carcinoma (OSCC) is the most common head and neck malignancy with ~30,000 patients diagnosed annually in North America. Despite advances in surgical, and chemo-radiation techniques, only ~50% of OSCC patients survive 5 years after diagnosis. Thus, there is an urgent clinical need to identify new targets for OSCC treatment.

Regulation of mRNA translation plays a critical role in the survival, proliferation, invasion and migration of cancer cells, and tumor angiogenesis. The non-canonical translation of distinct mRNAs has been implicated in several biological events leading to oncogenesis and cancer progression. Eukaryotic initiation factor 5B (eIF5B) is a key factor that drives the internal ribosome entry site (IRES)-mediated translation of distinct anti-apoptotic proteins. eIF5B is also implicated in the pathophysiology of several malignancies. Our data from patient-derived tissue microarray suggest that eIF5B expression is significantly higher in OSCC compared to tumor-adjacent normal tissue. Analysis of publicly available single-cell RNAseq data demonstrated that *EIF5B* is predominantly expressed in cancer cells compared to other cells in the OSCC microenvironment and increased *EIF5B* expression is associated with poor prognosis for OSCC patients. RNAi-mediated eIF5B depletion in OSCC cells significantly increased TNF-related apoptosis-inducing ligand (TRAIL), radiation, or cisplatin-induced cell death. eIF5B depletion in OSCC cells resulted in decreased mRNA translation of distinct anti-apoptotic proteins and angiogenic factors. Bromodeoxyuridine (BrdU) incorporation, Matrigel™ invasion, and wound healing assays suggested that eIF5B depletion also inhibits OSCC cell proliferation, invasion, and migration, respectively. Furthermore, eIF5B depletion blunted pro-growth, pro-inflammatory, and pro-angiogenic signaling pathways resulting in decreased endothelial tube formation. Combined with these data, our ongoing work on preclinical mouse model has strong potential to establish eIF5B as a therapeutic target for OSCC.

Leveraging large-scale datasets and single cell omics data to develop a polygenic score for cisplatin-induced ototoxicity

Deanne Nixie R. Miao[1], Mackenzie A. P. Wilke[1], John Pham[1], Feryal Ladha[1], Mansumeet Singh[1], Janilyn Arsenio[1], Emilia Luca[2], Alain Dabdoub[2], Wejian Yang[3], Jun J. Yang[3], Britt I. Drögemöller[1]

[1]University of Manitoba, Winnipeg, MB, Canada
[2]University of Toronto, Toronto, ON, Canada
[3]St. Jude Children's Research Hospital, Memphis, TN, USA

Presenting author: britt.drogemoller@umanitoba.ca

Keywords: Cisplatin-induced ototoxicity, polygenic score, single cell RNA sequencing, precision medicine, pharmacogenomics

Background: Cisplatin-induced ototoxicity (CIO), characterized by irreversible and progressive bilateral hearing loss, is a prevalent adverse effect of cisplatin chemotherapy. Genome-wide association studies (GWAS) have highlighted the polygenicity of this adverse drug reaction. Therefore, this study aimed to develop a polygenic score for CIO using large-scale hearing loss GWAS data combined with inner ear single cell multi-omics data.

Methods: We developed a hearing loss polygenic score (PGS_{HL}) using UKBiobank data (n=353,983) and SBayesR. To enhance the relevance of this score to CIO (PGS_{CIO}), we used murine inner ear single nuclei RNA sequencing (snRNA-seq) data and Milo-R to identify cells that exhibited differential abundance post-cisplatin treatment, and selectively included variants mapping to differentially expressed genes. We tested the association between the two PGSs and CIO in two independent CIO cohorts: the PanCareLIFE (PCL) cohort (n=390) and the St. Jude Children's Research Hospital (SJMB) cohort (n=238), using ReACt and logistic regression.

Results: PGS_{CIO} (P=5.5x10^{-5}, R^2=0.04) outperformed PGS_{HL} (P=2.9x10^{-3}, R^2=0.023) in predicting CIO in the PCL cohort. PGS_{CIO} was also associated with CIO in the SJMB cohort (P=0.04, R^2=0.02). Analysis of the scRNA-seq data generated through this study identified cisplatin-induced changes in the proportion of specific cochlear cell types, including specialized auditory cells, supporting cells, immune cells and bone cells.

Conclusion: We developed the first PGS for CIO using a large-scale hearing loss dataset and a biologically-informed filter from inner ear snRNA-seq data. Our approach offers new avenues for developing pharmacogenomic PGSs, and identified cochlear cells that may play critical roles in CIO. We are currently refining this novel genomic prediction model through the inclusion of snATAC-seq data, alongside the snRNA-seq data.

Multi-omic Dissection of Lymph Node Metastasis in Oral Squamous Cell Carcinoma

Ayan Chanda[1], Daniel Young[1,6], Steven C. Nakoneshny[2], T. Wayne Matthews[2,4], Shamir Chandarana[2,4], Robert D. Hart[2,4], Joseph C. Dort[2,4,5], Antoine Dufour[1,6], Martin D. Hyrcza[2,3], Pinaki Bose[1,2,3,5].

[1]Department of Biochemistry and Molecular Biology, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada.

[2]Ohlson Research Initiative, Arnie Charbonneau Cancer Institute, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada.

[3]Department of Pathology and Laboratory Medicine, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada.

[4]Department of Surgery, Section of Otolaryngology – Head & Neck Surgery, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada.

[5]Department of Oncology, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada.

[6]Department of Physiology and Pharmacology, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada.

Presenting author: achanda@ucalgary.ca

Keywords: Oral squamous cell carcinoma (OSCC), Lymph node metastasis (LNM), Multi-omics, Metastatic drivers, Biomarker discovery, Non-coding alterations.

Lymph node metastasis (LNM) is a critical prognostic factor in oral squamous cell carcinoma (OSCC), yet the molecular events underpinning this process remain insufficiently understood (1). To investigate the biological mechanisms driving LNM, we performed a comprehensive multi-omic analysis of 60 primary OSCC tumors with matched adjacent normal tissues, stratified by nodal involvement. This study is a part of the Marathon of Hope Cancer Centres Network (MOHCCN) Gold Cohort, led by the Terry Fox Research Institute (2). Each sample was profiled using whole genome sequencing (WGS), whole transcriptome sequencing (WTSeq), Assay for Transposase Accessible Chromatin by sequencing (ATAC-seq) for chromatin accessibility, mass spectrometry-based proteomics, and metabolomics screen, generating a layered view of muti-omic variation.

This dataset enabled the integration of diverse molecular features to uncover coordinated alterations linked to metastatic progression. Leveraging multi-omic factor analysis (MOFA2) (3), we identified latent factors that distinguish tumors with LNM from those without, capturing complex interactions between omics layers. These factors revealed underlying regulatory programs and variation patterns, in primary tumors, not apparent within individual data modalities, highlighting the value of unsupervised integrative approaches in revealing hidden biological structure.

In parallel, recurrent non-coding somatic mutations were identified in LNM-positive tumors. These variants are enriched in regions outside canonical coding sequences, suggesting potential roles in

transcriptional regulation. Ongoing efforts are focused on intersecting these non-coding sites with regions of altered chromatin accessibility to evaluate their potential functional impact.

By combining large-scale data across molecular domains, our study aims to define robust, biologically informed signatures of LNM in OSCC. This integrative framework not only supports mechanistic hypothesis generation but also holds promise for the identification of biomarkers relevant to risk stratification and therapeutic targeting. Continued analysis will extend these findings through spatial transcriptomics and external validation cohorts.

This work underscores the importance of systems-level approaches in oncology, where multifactorial processes such as metastasis are best understood through a composite lens of genomics, epigenomics, and proteomics.

REFERENCES

1. Irani S. New Insights into Oral Cancer-Risk Factors and Prevention: A Review of Literature. Int J Prev Med. 2020;11:202.

2. Terry Fox Research Institute Marathon of Hope Cancer Centers Network. Electronic address mbc, Terry Fox Research Institute Marathon of Hope Cancer Centers N. The Terry Fox Research Institute Marathon of Hope Cancer Centres Network: A pan-Canadian precision oncology initiative. Cancer Cell. 2025;43(4):587-92.

3. Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni JC, et al. Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. Mol Syst Biol. 2018;14(6):e8124.

Bioinformatics as an Essential Component of a Biology Undergraduate Degree

Elisabeth Richardson[1] and Sara J. Smith[1]

[1] Department of Biology, Mount Royal University, Calgary, Alberta

Presenting authors: erichardson@mtroyal.ca and ssmith6@mtroyal.ca

Keywords: bioinformatics, scholarship of teaching and learning, pedagogy, education, computer science

Bioinformatics is a rapidly growing field at the intersection of biology and computer science, aimed at analyzing and interpreting complex biological data. However, neither computer science nor bioinformatics are consistently included in undergraduate biology curricula in Canada. There is a widespread assumption that, as "digital natives", the current cohort of undergraduates will enter the university with an advanced understanding of technology and computer science, but this does not appear to be reflected in assessments of digital literacy or computational skills. Advances in bioinformatics and the increasing prevalence of "big data" in research has exposed this gap in undergraduate education.

In this condensed version of our keynote presentation from the Western Canadian Conference for Computing Education, we discuss challenges and opportunities of integrating coding, data science, and bioinformatics into a standard biology undergraduate curriculum based on our experiences as professors at Mount Royal University, an undergraduate institution. We emphasise the importance of including basic data science, data literacy principles, and process skills as early as possible, and practical guidance for building student confidence in their computational and data science skills. We also propose how resources like the Canadian Bioinformatics Hub can be applied to undergraduate education, and a framework for strengthening routes into bioinformatics training programmes, graduate degrees, and industry careers in Alberta.

The role of groundwater microbiota in the natural bioremediation of nitrate

Madison Laing[1], Gama Cabria[1], Pauline Humez[1], Michael Nightingale[1], Bernhard Mayer[1], Jean Birks[2], Marc Strous[1]

[1] Department of Earth, Energy, and Environment, University of Calgary

[2] Alberta Environment and Protected Areas, Calgary

Southern Alberta is characterized by intense agricultural activity, hosting 40% of the cattle and calf population in Canada and applying nearly 75% of all irrigation water across Canada. Alberta

is also a leader in grain production and ranks as the second-highest user of commercial fertilizers. Use of manure and synthetic fertilizers could contaminate groundwater, which is used as drinking water by many households. The maximum acceptable concentration for nitrate is 10 mg-N/L.

Analysis of nitrate in >3,500 groundwater samples from Southern Alberta shows this limit is exceeded in 10% of cases. Although broad, general trends implicate manure as a source of nitrate in groundwater, nitrate concentrations are highly uneven. Even in areas of intense manure production and fertilizer application, groundwater frequently contains low or undetectable nitrate. Denitrification, the bacterial conversion of nitrate into harmless nitrogen gas, could explain the absence of nitrate in many groundwater samples. Here, we use shotgun sequencing of groundwater microbiomes to develop a denitrification-index. Our data show that 16S rRNA gene amplicon sequencing is not useful to assess denitrification; analysis of metagenome assembled genomes shows that strains of the same species differ in their ability to denitrify. We also incubate groundwater in mesocosms with different rock types and find that rock type strongly determines denitrification rate. Denitrification requires an electron donor, usually organic compounds, reduced sulfur or iron. This electron donor is sourced from the rocks or sediments surrounding the groundwater. Our experiments suggest that the supply of electron donors from rocks or sediments may eventually limit the natural attenuation of nitrate in groundwater.

The origin and spread of *Culex pipiens* in Alberta

Huiqing Yeo[1], John Soghigian[1]

[1] University of Calgary, Calgary, Canada

Presenting author: huiqing.yeo@ucalgary.ca

Keywords: Population genomics, invasion biology, mosquito, vectors, West Nile virus

Culex pipiens is an invasive species in Canada capable of transmitting many pathogens of human and veterinary concern, including West Nile virus, St. Louis encephalitis virus, and avian malaria, all of which are present concerns in North America [1]. Within Canada, *Culex pipiens*'s distribution has been largely limited to eastern Ontario, Quebec, the Maritimes, and British Columbia, but global climate change is expected to result in significant changes to its distribution [2]. In fact, we have already detected *Culex pipiens* in Alberta, well ahead of the predicted timeline – Edmonton in 2018, and Calgary in 2022 [3], highlighting its rapid spread into new areas and its status as a considerable invasive vector threat. However, despite its public health importance and clear evidence of potential as a significant disease vector in most of its range [4,5], relatively little is known about the public health risk this mosquito may pose in areas it has recently invaded in Canada.

We carried out whole genome sequencing on *Culex pipiens* specimens from Calgary, Edmonton and Red Deer, and supplemented the dataset with genomes from Genbank and collaborators encompassing distributions from USA, Belarus, Kyrgyzstan and Russia. We found that *Culex pipiens* populations across Alberta likely originated from one introduction event and surprisingly exhibits a population structure that is genetically similar to eastern US populations (e.g., Maine). These results will be important to infer invasion pathways and effectively identify target sites and dispersal routes for *Culex pipiens* vector control.

REFERENCES

1. https://doi.org/10.4269/ajtmh.21-0167
2. https://doi.org/10.1016/j.apgeog.2011.05.015
3. https://doi.org/10.1111/mve.12408
4. https://doi.org/10.1126/science.1094247
5. https://doi.org/10.1016/j.meegid.2011.08.013

Optimizing Viral Detection in Wastewater Metatranscriptomes Through Targeted Depletion of Tobamoviruses and Ribosomal RNAs

Maria A. Bautista[1], Tulika Bhardwaj[1], Dylan J. Barbera[2], Samuel J. Rulli[2], Janine McCalder[1], Chloe Papparis[1], Xiaoli Pang[3], Bonita Lee[3], Rhonda G. Clark[1], Christine O' Grady[4], Michael D. Parkins[5,6] and Casey R.J. Hubert[1]

[1] Department of Biological Sciences, University of Calgary, Calgary, Canada
[2] QIAGEN Sciences Inc, Frederick, MD, USA
[3] Department of Laboratory Medicine & Pathology, Faculty of Medicine & Dentistry, University of Alberta, Edmonton, Canada
[4] Advancing Canadian Water Assets, Calgary, Canada
[5] Department of Microbiology, Immunology, and Infectious Diseases, University of Calgary, Calgary, Canada
[6] Alberta Health Services, Calgary, Canada

Presenting author: tulika.bhardwaj@ucalgary.ca

Keywords: Wastewater, rRNA-depletion, RiboDetector, tobamovirus, virome

Wastewater is dynamic and contains a complex biological mixture of nucleic acids from many sources, including humans, associated microbiota from humans, and dietary residues. Depending on the sewage infrastructure, industrial and agricultural runoff can also be present. This diverse genetic milieu presents opportunities for metatranscriptomic analysis to track microbial community dynamics including the presence and prevalence of pathogens and emerging viral threats. However, wastewater metatranscriptomes can be overwhelmed by extremely abundant ribosomal RNAs (rRNAs) and RNA genomes from plant-infecting tobamoviruses derived from human diets. These RNA signals make it extremely challenging to detect viral species present at lower concentrations, potentially limiting the sensitivity of pathogen surveillance efforts. To improve the representation of target viral reads, an enhanced depletion strategy was developed utilizing rRNA-targeting probes (QIAseq FastSelect 5S/16S/23S) and custom Tobamovirus RNA probes that was applied to wastewater samples collected from a wastewater treatment plant in Calgary, Alberta. We subsequently implemented a computational pipeline for metatranscriptomic analysis, integrating sequential data processing utilizing the bioinformatics tools RiboDetector [1], Kraken2 [2], geNomad [3], and CheckV [4]. As a result, the proportion of viral reads, particularly from the Riboviria realm, increased by 90%. Despite these improvements, overall viral read counts remained low, underscoring the need for further refinements in depletion strategies to minimize bacterial nucleic acid (rRNA) signal and enrich for low-abundance RNA viruses. This study highlights the potential of targeted depletion approaches to enhance the sensitivity and cost-effectiveness of WBS for viral pathogen monitoring in wastewater.

REFERENCES

1. https://doi: 10.1093/nar/gkac
2. https://doi.org/10.1186/s13059-019-1891-0
3. https://doi.org/10.1038/s41587-023-01953-y

4. https://doi.org/10.1038/s41587-020-00774-7

Leveraging Artificial Intelligence and the Cloud in Advancing Agricultural Crop Research

Nathaniel Zhin-Loong Lim[1], Harpreet Bargota[2], Tobey Haonan Wang[1], David Enoma[3], Mathew Richards[2], Kamaljeet Gill[1], Manoj Natarajan[1], Trista Cormier[4], Vybhavi Gururaj[2], Lance Lansing[1], Arun Kommadath[5], Elie Raherison[6], Deng-jin Bing[5], Kimberley Burton Hughes[1], Rajiv Tripathi[5], Ian Stavness[7], Quan Long[3], Christine Lowe[8], Jackson Eyres[2], Étienne Low-Décarie[2], Stacy Singer[1], Keshav Singh[1], Nicholas Larkan[4], Ravinder Goyal[5], Xiaohui Yang[2], Rodrigo Ortega Polo[1]

[1] Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada (AAFC), Lethbridge AB, Canada
[2] Biological Informatics Centre of Excellence, Agriculture and Agri-Food Canada (AAFC), Ottawa ON, Canada
[3] University of Calgary, Calgary AB, Canada
[4] Saskatoon Research and Development Centre, Agriculture and Agri-Food Canada (AAFC), Saskatoon SK, Canada
[5] Lacombe Research and Development Centre, Agriculture and Agri-Food Canada (AAFC), Lacombe AB, Canada
[6] Swift Current Research and Development Centre, Agriculture and Agri-Food Canada (AAFC), Swift Current SK, Canada
[7] University of Saskatchewan, Saskatoon SK, Canada
[8] Digital Solutions and Innovation, Agriculture and Agri-Food Canada (AAFC), Ottawa ON, Canada

Presenting author: NathanielZhin-Loong.Lim@AGR.GC.CA
Corresponding author: Rodrigo.OrtegaPolo@AGR.GC.CA

Keywords: WGRF, AAFC, Agriculture, Genomics, Phenomics, Artificial Intelligence, Cloud, SegmentAnything, IMAS

In the agricultural sector, the constant development of better crop cultivars is necessary for us to address the pressures of a growing human population and climate change. Current crop breeding programs routinely rely on classical approaches and genomics technologies. The development of superior crops can be further accelerated with the adoption of game-changing technologies such as artificial intelligence and cloud-computing; both of which have demonstrated utility in biomedical research.

With funding support from the Western Grains Research Foundation, we initiated a project to investigate the benefits of adopting artificial intelligence and cloud-computing technologies to enhance the analyses of omics data of agricultural crops (alfalfa, faba bean, and field pea). This project has multiple aims, including (1) the creation and utilization of reproducible and streamlined workflows for the analyses of genomics and phenomics data; (2) implementation of phenomics pipelines on cloud platforms to evaluate the scalability and burst capacity of said platforms; and (3) the employment of advanced analytical methods to integrate sequenced-based (i.e., Genotype-By-Sequencing) omics data with image-based phenomics data to enhance predictive analytics for plant breeding.

In our presentation, we will share some insights from this project including (1) leveraging the SegmentAnything[1] model for extracting image-based phenomics data for faba beans on the cloud, (2) assessing multiple phenomics data modalities for detecting water-starvation stress in alfalfa plants, and

(3) utilizing Image-Mediated Association Studies “IMAS”[2] techniques to identify candidate genes of interest for a faba bean phenotype.

REFERENCES

1. https://doi.org/10.48550/arXiv.2408.00714
2. https://doi.org/10.1016/j.ajhg.2023.11.006

Beyond SNPs: Developing a structural variant detection workflow and associated database for Holstein cattle

Emily K. Herman[1], Jason R. Grant[1], Lael D. Barlow[1], Filippo Miglior[2,3], Flavio S. Schenkel[2], Christine F. Baes[2,4], Paul Stothard[1]

[1]Agricultural, Food & Nutritional Science, University of Alberta, Edmonton, Canada.
[2]Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, Canada.
[3]Lactanet, Guelph, Canada.
[4]Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, Switzerland.

Presenting author: eherman@ualberta.ca

Keywords: Livestock, Genomics, Structural variation, Cattle, Workflow, Snakemake

Structural variation is generally defined as chromosomal genomic rearrangements greater than 50 bp, including changes in copy number, orientation, and chromosomal translocations and fusions. Structural variants (SVs) are an important source of monogenic and complex trait variation in cattle, although it is less explored than single-nucleotide variation.

We developed a structural variant calling workflow using two popular SV callers (Manta and Smoove), testing multiple SV caller programs with the Genome in a Bottle benchmark structural variant callset from NCBI. We then used the workflow and publicly available Illumina whole genome sequence data from 310 Holstein cattle to produce a large and deeply characterized collection of duplication and deletion variants. SV genotypes from each caller were confirmed to accurately recapitulate animal relationships estimated using WGS SNP genotypes from the same dataset, with Manta genotypes outperforming Smoove, and deletions outperforming duplications. These 31K (Manta) and 68K (Smoove) SVs, along with extensive annotations, gene feature information, and BAM file content, are available as part of a custom interactive database (https://svdb-dc.pslab.ca).

Our analysis identified SVs affecting genes such as *POPDC3, ORM1, G2E3, FANCI, TFB1M, FOXC2, N4BP2, GSTA3*, and *COPA*, showing how this database can be used to find well-supported genic SVs, determine SV breakpoints, design genotyping approaches, and identify processed pseudogenes masquerading as deletions. While this resource is geared towards exploring sequence variation in Holstein cattle, the workflow can be applied generally for SV discovery in other animals.

Decoding poultry threats: unveiling sequencing types, serogroups, and antimicrobial resistance of avian pathogenic *E. coli* in Alberta, Canada

Yanqi Wang[1], Mawra Gohar[1], Beverly Morrison[1], Frank Marshall[2], Teryn Girard3, Matthew Waldner[4], Matthew Walker[5], Trevor Alexander[5], Maud de Lagarde[6], M. Faizal Abdul Careem[1], John Morris Fairbrother[6], Dongyan Xu Niu[1]

[1] Faculty of Veterinary Medicine, University of Calgary, Calgary, Alberta, Canada
[2] Marshall Swine & Poultry Health Services, Camrose, Alberta, Canada
[3] Prairie Livestock Veterinarians, Red Deer, Alberta, Canada
[4] Public Health Agency of Canada, Winnipeg, Manitoba, Canada
[5] Agriculture and Agri-Food Canada, Lethbridge Research and Development Centre, Lethbridge, Alberta, Canada
[6] Department of Pathology and Microbiology, Faculty of Veterinary Medicine, Université de Montréal, Québec, Québec, Canada

Presenting and corresponding author: dongyan.niu@ucalgary.ca

Keywords: Avian pathogenic *Escherichia coli* (APEC), whole genome sequencing, multilocus sequence typing, antimicrobial susceptibility, serotypes

Background: Avian pathogenic *Escherichia coli* (APEC) is the causative agent of avian colibacillosis, leading to substantial economic losses in Canada and other regions. Avian fecal *E. coli* (AFEC) is considered a potential reservoir for APEC, due to shared genetic characteristics.

Hypothesis: Avian Pathogenic *Escherichia coli* (APEC) isolates from poultry in Alberta have distinct phenotypic and genotypic profiles, characterized by variations in virulence factors, antimicrobial resistance genes, and phylogenetic relatedness, which may contribute to differences in pathogenicity and adaptability to specific host and environmental conditions.

Objective: This study aimed to identify and profile phenotypic and genotypic characteristics of potential pathogenic *E. coli* that cause poultry colibacillosis in Alberta.

Methods and Results: A total of 100 APEC isolates from diseased and dead chickens (broilers and layers) and turkeys from 45 farms collected between 2021 and 2023 were subjected to serogroup typing, APEC-associated virulence gene (VG) profiling, antimicrobial susceptibility testing, whole genome sequencing, multilocus sequence typing, and phylogenetic analysis. Additionally, 121 AFEC isolates from five-layer flocks were collected to investigate APEC-associated VG carriage in healthy birds. Furthermore, genetic similarity between 22 AFEC strains identified as potential APEC (poAPEC) and APEC strains was assessed through phylogenetic analysis. A phylogenetic tree of 100 APEC isolates and 89 AFEC isolates was also constructed. We identified 27 serogroups, 38 known and 2 novel sequence types (STs), and varied VG distribution across APEC isolates from chickens and turkeys. Interestingly, there were variations in VG among AFEC strains across distinct egg laying stages or housing systems. Antimicrobial resistance (AMR) genes, especially extended-spectrum beta-lactamase genes, were detected in APEC and poAPEC strains across all poultry species. Moreover, phylogenetic analysis revealed close clonal relationships among APEC and poAPEC strains. Notably, we identified multiple fecal isolates that shared identical ST and

serotypes with APEC strains. This study is the first to document characteristics of APEC and poAPEC isolates that cause or potentially cause colibacillosis morbidity and mortality in chickens and turkeys in Alberta. A phylogenetic tree of 20 APEC isolates from Québec with representative AFEC, APEC and pAPEC was also constructed, indicating genetic relatedness with APEC, AFEC and poAPEC.

Conclusions: Various APEC phenotypes and genotypes can lead to colibacillosis in poultry, including layers and may originate from commensal E. coli present in the gastrointestinal tract of birds.

A genomic admixture testing method for bison conservation

Lael D. Barlow[1] and Paul Stothard[1]

[1] University of Alberta, Edmonton, Canada

Presenting author: lael@ualberta.ca

Keywords: Admixture, Ancestry, Wildlife, Bison, Cattle, Livestock, Breeding, Whole-genome sequencing

The long-term survival of wild bison in North America is under threat, as surviving herds constitute only a tiny fraction of historical population sizes. Planned conservation efforts include strategic transfer of animals and germplasm among small isolated herds to prevent loss of genetic diversity. A complicating factor is that, due to historical mismanagement, some wild bison have varying degrees of mixed ancestry between the Wood and Plains bison subspecies and domestic cattle. Hence, to support the strategic transfer of genetics among herds, accurate tests are needed to determine the extent to which individual bison are an admixture of these source populations. We developed a method for this based on the latest available whole genome sequencing data for modern bison and neohistoric bison, which are free of admixture. Results for real and simulated data indicate that estimated admixture proportions are sufficiently accurate to provide reliable guidance for conservation actions. Also, the developed software offers a unique combination of features, making it readily applicable to other species.

Clustering in single-cell data using Gene-Gene Correlation Structure

Sandesh Acharya[1,3,4,5], Dinghao Wang[2], Jiami Guo[1,3,4,5,6], Qingrun Zhang[2,3,4,7,8]

[1]Department of Cell Biology and Anatomy, University of Calgary, Calgary, AB, Canada.
[2]Department of Mathematics and Statistics, University of Calgary, Calgary, AB, Canada.
[3]Alberta Children's Hospital Research Institute, University of Calgary, Calgary, AB, Canada.
[4]Hotchkiss Brain Institute, University of Calgary, Calgary, AB, Canada.
[5]Department of Medical Science, University of Calgary, Calgary, AB, Canada
[6]New York Stem Cell Foundation-Robertson Investigator, New York, USA.
[7]Department of Biochemistry and Molecular Biology, University of Calgary, Calgary, AB, Canada
[8]Data Science Advisory Unit, University of Calgary, Calgary, AB, Canada

Presenting Author: sandesh.acharya@ucalgary.ca

Keywords: Single-cell RNA sequencing, Clustering, Gene-Gene Correlation, Dendrogram Clustering, Biclustering, WGCNA.

Genes do not function in isolation, they form intricate networks within cells, influencing various cellular processes like differentiation, development, and response to stimuli. This interplay involves the coordinated regulation of multiple genes which contributes to the establishment or maintenance of distinct cell types. We propose that each cell type can be characterized by a set of genes which exhibit high correlations with one another.

Building upon this premise, we have developed a novel bi-clustering algorithm tailored for single-cell data analysis, leveraging gene-gene correlation patterns. Our algorithm iteratively identifies clusters of genes exhibiting strong mutual correlations and segregates cells based on the expression profiles of these gene networks. In the first step, a network of highly correlated genes is identified using a simplified WGCNA algorithm [1]. Next, the dataset is clustered into two groups by using the subset of genes and K-means algorithm [2]. For each of the sub-clusters identified, we iteratively run both steps until there are fewer than 50 cells in a cluster or no correlated genes. As a result, this method outputs a dendrogram showing the relations between parents and daughter clusters as well the genes that can separate them.

To validate the efficacy of our approach, we have applied our algorithm to four distinct single-cell RNAseq datasets [3,4,5]. Our results demonstrate the algorithm's ability to effectively segregate the datasets into biologically meaningful clusters. Notably, the genes identified at each iterative step serve as reliable discriminators, facilitating the accurate distinction between the clusters.

1. https://doi.org/10.1186/1471-2105-9-559
2. https://doi.org/10.2307/2346830
3. https://doi.org/10.1371/journal.pcbi.1006361
4. https://doi.org/10.1016/j.cell.2016.07.054
5. https://doi.org/10.1016/j.cels.2016.08.011

asmR package in R for nanopore adaptive sampling bacterial metagenomic data analysis

Anatoliy Trokhymchuk[1], Cheryl Waldner[1], Matthew Links[3]

[1]Department of Large Animal Clinical Sciences, Western College of Veterinary Medicine, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
[2]Department of Animal and Poultry Science, College of Agriculture and Bioresources, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
[3]Department of Computer Science, College of Arts and Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

Presenting author: **anatoliy.trokhymchuk@usask.ca**

Keywords: Native DNA sequencing, nanopore sequencing, nanopore adaptive sampling, bacterial metagenomics, composite data, information theory, R

One of the proposed fourth-generation biopolymer sequencing platforms definitions includes obtaining native data from a sample without an amplification bias. This approach results in an incredibly rich raw data generated by sequencing experiments. Raw reads generated with the pioneer in fourth-generation domain Oxford Nanopore Technologies (ONT) platform vary in length from tens to hundred million bases with quality ranging between Q9 to Q40. While potentially tremendously beneficial in resolving biological and clinical questions, this type of data presents several major challenges for bioinformatic analysis. Traditional bioinformatic analysis concepts have evolved with previous generations of sequencing platforms and are bound by their intrinsic limitations. Legacy bioinformatics might not be applicable or helpful with answering questions enabled by the unprecedented high dimensionality level of the fourth-generation sequencing data analysis. While it is possible to use established bioinformatic analysis tools and pipelines for nanopore data analysis after consensus-based assembly and error correction, raw reads generated from native DNA libraries carry unique information that can be uncovered by bespoke analyses and deployed for practical applications. For example, nanopore adaptive sampling (NAS) is a unique technological capability of the ONT platform based on rapid raw reads mapping against a provided reference and selective strands rejection in close-to-real-time (1, 2). Under favorable conditions, NAS can result in up to 30-fold on-target data yield increase. Sequencing data generated by ONT NAS carry additional information of reads status assigned during the run (“continue receiving”, “stop receiving”, and “reject”). Bioinformatic tools for NAS raw reads data rapid analysis and NAS performance metrics need to be established to support the emerging bacterial metagenomic applications (3-5). The “Adaptive sampling tools in R” (*asmr*) package is built to perform ONT NAS metagenomics data analysis respecting the compositional nature of the input dataset and not relying on the convention of reads consensus building (6, 7). Using an example application of Bovine Respiratory Diseases (BRD) metagenomics-based diagnostic testing method development, *asmr* provides a seamless functionality for sampling effort equilibration, reads classification, grouping by a list of references and NAS status, cleaning, aggregating, and graphical presentation. It allows intuitive numerical, statistical, and visual evaluation of the NAS effects on the experiment outcomes.

1. https://doi.org/10.1093/bioinformatics/bty191
2. https://www.nature.com/articles/nmeth.3930.pdf
3. https://doi.org/10.1093/bib/bbac453
4. https://doi.org/10.1186/s13059-021-02582-x
5. https://www.biorxiv.org/content/biorxiv/early/2021/09/10/2021.08.29.458107.full.pdf

6. https://doi.org/10.1038/s41579-018-0029-9
7. https://pubmed.ncbi.nlm.nih.gov/30929407/

Acclimation of *Escherichia coli* at sublethal concentrations of antimicrobial coinage metals reveals novel tolerance mechanisms through RNA-seq

Daniel A. Salazar-Alemán[1], Ashley McGibbon[1], Raymond J. Turner[1]

[1]Department of Biological Sciences, Faculty of Science, University of Calgary, Calgary, Alberta, Canada.

Presenting author: danielarmando.salaza@ucalgary.ca

Keywords: iron metabolism, metal resistance, stress adaptation, transcriptomics, thiol metabolism, protein misfolding, envelope stress, copper resistance

In the current antimicrobial resistance era, MBAs have re-emerged as an alternative to manage bacterial infections. Silver, copper and gold have drawn interest due to their antimicrobial applications[1]. Despite the effectiveness of these coinage metals, their exact mechanisms of action remain a knowledge gap. Additionally, metal-induced selection has been demonstrated to co-select for resistance to organic antibiotic compounds[2]. This is a looming threat, particularly in the context of sublethal exposures that can lead to resistance. For this reason, we present a full systems overview of a culture that has acclimated to grow in the presence of a sublethal inhibitory concentration of coinage metals. By interrogating the *Escherichia coli* transcriptional regulatory network, we identified and compared biological processes needed for fitness when growing in planktonic form while being exposed to sublethal inhibitory concentrations of silver nitrate ($AgNO_3$), copper sulfate ($CuSO_4$) or tetrachloroauric acid ($HAuCl_4$). To achieve this, *E. coli* K12 BW25113 was cultured in M9-glucose media for 10 hours at 37 ºC 150 rpm. $AgNO_3$ 7 µM, $CuSO_4$ 39 µM and $HAuCl_4$ 10 µM for were identified as ideal concentrations that would challenge growth but not appreciably reduce cell yield for sampling. Differential gene expression was determined by contrasting metal treatments against a non-metal treated control; significance parameters were set at average expression fold-change > 2 and adjusted p- < 0.05. Silver, copper and gold yielded 131, 80 and 1028 significantly up-regulated genes. 23 of those were shared across the three conditions, with six of them related to maintaining iron homeostasis. This, combined with the high expression activity presented in the Fur regulon, points to a dysregulation of iron homeostasis upon prolonged metal exposure. Induction of the copper homeostasis systems Cue and Cus in our copper experiment was expected, but surprisingly gold increased the expression of Cue, while silver did not register significant changes. The CpxR regulon was up-regulated only in our silver treatment, a sign of protein misfolding at the cell envelope level. CysB is a transcription factor that modulates cysteine and sulfur metabolism, and its target genes had a high activity in the copper and gold treatments. Overall, this work establishes a precedent for common and unique patterns of expression after prolonged metal stress. This is a significant step in understanding how bacteria can adjust their physiology to coexist with sublethal concentrations of these antimicrobial group IB metals according to their different mechanisms of action and relative toxicity.

1. https://doi.org/10.1007/978-3-030-97185-4_4
2. https://doi.org/10.1111/jam.15492

Identifying functional non-coding variants in glioblastoma evolution

Ana Nikolic [1], Sorana Morrissy[2], Jennifer Chan[3], Amy Banks[4]

[1] University of Calgary, Calgary, Canada
[2] University of Calgary, Calgary, Canada
[3] University of Calgary, Calgary, Canada
[4] University of Calgary, Calgary, Canada

Presenting author: amy.banks@ucalgary.ca

Keywords: Epigenetics, glioblastoma, single-nucleotide variant, oncology, non-coding genome, regulatory plexus

Glioblastoma is the most common form of primary malignant brain tumour in adults, with a median survival of just under 15 months, despite aggressive treatment1. These tumours have high levels of intratumoral heterogeneity, with a population of self-renewing glioma stem cells, as well as multiple tumour cellular states [2,3]. Despite an increased understanding of genetics and heterogeneity in this disease, treatment plans have changed little in the last decade, and many discoveries, including cell states, have not been integrated into clinical practice [4]. Through investigating the non-coding genome of this disease, the influence of mutations in enhancer elements of the epigenome will be explored, with the aim of expanding our understanding of the mechanisms behind glioblastoma and highlight potential treatment avenues.

The goal of this project is to identify functional non-coding variants influencing glioblastoma disease progression. Assay for Transposase Accessible Chromatin with sequencing (ATAC-seq) is a tool used to map regions of the whole genome which are in an open, euchromatic state, therefore accessible to transcription factors and interaction with regulatory elements [5]. This technique is leveraged to filter the non-coding region by regulatory significance, and has been carried out on glioblastoma samples to describe the epigenome of these tumours. Variants are called on these accessible regions by using a combination of multiple variant callers (Mutect2, Strelka2, DeepVAR) and their potential role in disease progression is examined. Regions significantly enriched for mutations will be identified through statistical tests, such as permutation and Hotspot of cis-Regulatory, Significantly-mutated Elements (HoRSE) analysis6, and annotated based on regulatory role by using bioinformatic tools, such as HOMER, and cis-regulatory element databases (CREdb, ENCODE). ATAC-seq was also done on healthy brain samples to generate a control dataset with the purpose of filtering variants and regions of high mutation rate common in healthy populations. Variants of functional significance are evaluated through various software tools, such as variant effects predictors (CADD, DeepSEA/SEI) and variant annotation tools (ANNOVAR). The variant annotations will be combined with the region annotations to identify non-coding functional variants associated with glioblastoma disease progression. Variants of significant relevance identified through bioinformatics tools will be functionally validated in a wet-lab setting through CRISPR screens to verify their role in glioblastoma progression. This project will provide valuable understanding to the role of the non-coding DNA of glioblastoma genomes and help to identify potential epigenetic targets to allow for development of further treatments.

1. https://doi.org/10.3390/cancers14102412
2. https://doi.org/10.1038/s41392-021-00491-w

3. https://doi.org/10.1016/j.cell.2019.06.024
4. https://doi.org/10.1200/OP.22.00476
5. https://doi.org/10.1038/s41467-021-23922-2
6. https://doi.org/10.1158/1541-7786.MCR-2

Sabrin Bashar[1], Hein M Tun[2], Joseph Ting[1], Matthew Hicks[1], Piush Kumar Mandhane[1], Theo J. Moraes[3], Elinor Simons[4], Stuart Turvey[5], Padmaja Subbarao[3], James Scott[3], Anita Kozyrskyj[1]

[1] University of Alberta, Edmonton, Canada,
[2] The Chinese University of Hong Kong, Hong Kong, China,
[3]University of Toronto, Toronto, Canada, 4University of Manitoba, Winnipeg, Canada,
[5]University of British Columbia, Vancouver, Canada

Presenting author: sbashar@ualberta.ca

Keywords: Hospital length-of-stay, Infant, Gut Microbiota, 16S rRNA sequencing, Mode of Delivery, Atopic Sensitization

This study determined whether postnatal hospital length-of-stay (LOS) following different birth modes is associated with gut microbiota pathways in 3- and 12-months infants and predicts atopic sensitization at ages 1 and 3 years. The gut microbiota of 1313 infants from the Canadian Healthy Infant Longitudinal Development (CHILD) Cohort Study were analyzed through Illumina 16S rRNA sequencing of fecal samples collected at 3-and 12-months. The gut microbial profiles of infants hospitalized for >1 day in vaginal delivery (VD) and ≥3 days in cesarean delivery (CD) were compared with those of infants with shorter hospital stays. Associations between hospital LOS, gut microbiota composition, and atopic (food and inhalant) sensitization were evaluated using multivariable logistic regression and sequential mediation analyses. At 3 months, infants with prolonged hospitalization exhibited depletion of commensal microbes Bacteroides in VD infants (p=0.03) and Bifidobacterium in CD infants (p=0.025). In VD infants without maternal intrapartum antibiotic (IAP) exposure, a longer LOS was associated with higher abundance of pathobionts Enterococcus (aOR 1.41, 95%CI 1.04-1.93) and Citrobacter (aOR 1.42, 95%CI 1.04-1.94) at 3 months, and lower abundance of Bacteroidaceae (aOR 0.74, 95%CI 0.54-1.01) at 3 months and 12 months. Sequential mediation analysis revealed higher abundance of Enterococcus or Citrobacter at 3-months with Bacteroidaceae depletion at 12 months as a microbiota pathway from LOS to atopic sensitization at 1-year ($Effect_{Enterococcus}$ 0.0117 bootstrap 95%CI 0.0003, 0.0307 or $Effect_{Citrobacter}$ 0.0139 bootstrap 95%CI 0.0018, 0.0346) and food sensitization at 1-year ($Effect_{Citrobacter}$ 0.0141 bootstrap 95% CI 0.0016, 0.0355) and 3-years ($Effect_{Citrobacter}$ 0.0205 bootstrap 95% CI 0.0016, 0.0547) in VD-no IAP infants. No microbiota mediation paths were found for CD or VD infants exposed to maternal IAP. In conclusion, prolonged postnatal exposure to the hospital microbial environment, even without antimicrobial treatment, may lead to an overrepresentation of pathogenic bacteria and depletion of beneficial microbiota, ultimately contributing to atopic sensitization in later life.

Leveraging metagenomics to characterize the gut microbiome of woodland caribou for conservation.

Charlotte. Bourbon[1], Mason. R. Stothart[2], Margaret. M. Hugues[3], Helen Schwantje[4], Caeley Thacker[5], Marco. Musiani[6], Jocelyn. Poissant[1]

[1] Faculty of Veterinary Medicine, University of Calgary, Calgary, Alberta, Canada. https://orcid.org/0009-0004-5709-6995
[2] Department of Biology, University of Oxford, Oxford, United Kingdom
[3] Biological Sciences, University of Calgary, Calgary, Alberta, Canada
[4] British Columbia Ministry of Water, Land, and Resource Stewardship. Wildlife Branch. Nanaimo, BC, Canada. Emeritus
[5] British Columbia Ministry of Water, Land, and Resource Stewardship. Wildlife Branch. Nanaimo, BC, Canada
[6] Institute of Condensed Matter Chemistry and Technologies for Energy ICMATE - CNR, Corso Stati Uniti 4, Padova 35127, Italy

Presenting author: charlottebourbon@ucalgary.ca

Keywords: Caribou, metagenomics, shotgun sequencing, microbiome conservation, short reads

The gut microbiome provides a critical molecular lens for understanding wildlife health, adaptation, and conservation, particularly for endangered species confronting complex environmental challenges[1]. We conducted a comprehensive metagenomic investigation of the gut microbiome in caribou (*Rangifer tarandus*), an iconic species facing significant ecological pressures[2]. Using shallow shotgun sequencing, we analyzed 184 fecal samples from 21 distinct caribou herds across British Columbia's diverse landscapes, employing advanced bioinformatics to characterize bacterial, archaeal, and viral community structures. We showed that a minimum of ~500,000 paired end reads per sample could effectively reveal microbiome diversity, similarly found before on human[3] and feral horses[4]. We found that a conserved core microbiome emerged across herds, and there were pronounced structural variations between northern and southern herds. Notably, genetic factors demonstrated a more substantial impact on microbiome composition than geographical distance, with smaller, genetically isolated herds displaying markedly reduced microbiome diversity. Pathogen screening unveiled potential health risks, including the detection of *Listeria monocytogenes*, while functional metagenomic analyses showed remarkable metabolic redundancy. Of particular ecological significance, we identified adaptive mechanisms in microbial metabolic pathways critical for specific arboreal lichen digestion like starch degradation for southern herds. Notably, localized microbiome adaptations emerged, closely linked to distinct herds' foraging behaviors[5], highlighting the intricate relationship between diet, genetics, and microbial community structure. By integrating microbiome insights with conservation genomics, our research provides a data-driven framework for understanding caribou ecology and investigating population health. These findings offer novel, cost-effective methodological approaches for monitoring and supporting endangered wildlife, demonstrating the pivotal role of microbiome research in conservation biology.

1. West, A. G. *et al.* The microbiome in threatened species conservation. *Biol. Conserv.* 229, 85–98 (2019). https://doi.org/10.1016/j.biocon.2018.11.016

2. COSEWIC. *COSEWIC Assessment and Status Report on the Caribou Rangifer Tarandus, Northern Mountain Population, Central Mountain Population and Southern Mountain Population in Canada* (2014). https://www.registrelep sararegistry.gc.ca/virtual_sara/files/cosewic/sr_Caribou_Northern_Central_Southern_2014_e.pdf
3. Hillmann, B. *et al.* Evaluating the Information Content of Shallow Shotgun Metagenomics. *mSystems* 3, e00069-18 (2018). https://doi.org/10.1128/msystems.00069-18
4. Stothart, M. R., McLoughlin, P. D. & Poissant, J. Shallow shotgun sequencing of the microbiome recapitulates 16S amplicon results and provides functional insights. *Mol. Ecol. Resour.* 23, 549–564 (2023). https://doi.org/10.1111/1755-0998.13713
5. Webber, Q. M. R., Ferraro, K. M., Hendrix, J. G. & Vander Wal, E. What do caribou eat? A review of the literature on caribou diet. *Can. J. Zool.* 100, 197–207 (2022). https://doi.org/10.1139/cjz-2021-0162

Yuxuan Chen[1]

[1] University of Calgary, Calgary, Canada

Presenting author: yuxuan.chen1@ucalgary.ca

Keywords: single-cell sequencing, multi-omics, multiplex community detection, cell-to-cell networks, Louvain clustering

Clustering cells based on single-cell multi-modal sequencing technologies provides an unprecedented opportunity to create high-resolution cell atlas, reveal cellular critical states and study health and diseases. However, effectively integrating different sequencing data for cell clustering remains a challenging task. Motivated by the successful application of Louvain in scRNA-seq data, we propose a single-cell multi-modal Louvain clustering framework, called scMLC, to tackle this problem. scMLC builds multiplex single- and cross-modal cell-to-cell networks to capture modal-specific and consistent information between modalities, then adopts a robust multiplex community detection method to obtain the reliable cell clusters. In a comparison with fifteen state-of-the-art clustering methods on seven real datasets simultaneously measuring gene expression and chromatin accessibility, scMLC achieves better accuracy and stability in most datasets. Synthetic results also indicate that the cell-network-based integration strategy of multi-omics data is superior to other strategies in terms of generalization. Moreover, scMLC is flexible and can be extended to single-cell sequencing data with more than two modalities.

Screening anthelmintic drug resistance mutations in livestock parasites using Nanopore long-read sequencing

Rebecca Chen[1], Libby Redman[1], Lynsey Melville[2], Dave Bartley[2], John Gilleard[1]

[1] University of Calgary, Calgary, Canada
[2] Moredun Research Institute, Bush Loan, UK

Presenting author: pcrchen@ucalgary.ca

Keywords: Nanopore, resistance, anthelmintic, benzimidazole, β-tubulin

Routine anthelmintic drug use has resulted in widespread resistance in many parasitic nematode (roundworm) species, and it is crucial to minimize the development and dissemination of resistance to maximize the useful life of the few available drugs. Of the various anthelmintic drug classes, benzimidazole resistance (BZ-R) is the most well characterized. Single nucleotide polymorphisms (SNPs) at codons 167, 198, and 200 in the drug target, isotype-1 β-tubulin, have been shown to confer BZ-R. And new DNA sequencing technologies are revolutionizing our ability to undertake surveillance and study how drug resistance emerges and spreads. Several studies have started applying Illumina short-read amplicon sequencing to BZ drug target to screen for resistance-conferring mutations at a greater sensitivity and scale.

However, a major limitation in using short-read sequencing is that the sequenced fragment cannot be longer than 600 bp, which only covers a small region compared to the whole β-tubulin gene at around 3000-4000 bp. This limits our ability to detect other potential BZ-R conferring mutations elsewhere on the locus. Nanopore long-read sequencing have become increasingly affordable, but it's sequencing accuracy (1 error in 200-400 bases with the latest super accurate basecalling model) is still lagging behind compared to Illumina short-read sequencing platforms (1 error in 1000-10000 bases).

We have therefore designed a study to investigate the accuracy of BZ-R SNP detection using both Nanopore MinION long-read sequencing with results from Illumina MiSeq short-read sequencing of identical sample set as our ground truth. We chose to use Nematodirus battus, parasitic nematode of the sheep, collected from farms across the UK. This sample set was quite unique in that BZ-R was detected at an early stage, and many farms have low frequency of BZ-R F167Y SNP (1-5%) from Illumina sequencing results, allowing us to test the sensitivity and reliability of Nanopore sequencing on low frequency SNPs.

Nematodirus battus isotype-1 β-tubulin gene was amplified and sequenced using Nanopore MinION from 17 farm samples in 6 PCR replicates. A bioinformatic pipeline was developed to detect any SNPs resulting in non-synonymous mutation across the loci. We found the F167Y and F200Y SNP frequencies to be very consistent across the 6 replicates for each sample and the frequency was comparable to Illumina sequencing as well. We confirmed the viability of using Nanopore long-read sequencing for BZ-R SNP detection and have applied this method on several other sample sets to screen for potential BZ-R conferring mutations elsewhere on the locus.

Representation learning-based genome-wide association mapping improves gene discovery and disease risk prediction.

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

Keywords: Representation learning, Genome-wide association studies, Complex traits, Kernel Machine, Polygenic Risk Prediction.

Genome-wide association studies (GWAS) have provided key insights into the genetic basis of complex traits and diseases [1], [2]. However, traditional univariate approaches frequently fail to account for polygenicity, epistatic interactions, and linkage disequilibrium, resulting in missing heritability [3], [4] and diminished power to detect associations. We introduce Representation Learning-Based Association Mapping (RBAM), a novel framework that harnesses variational autoencoders (VAE) [5] to learn latent genotype representations. By reconstructing disease genotypes and incorporating encoder-decoder weights as variant-specific importance measures, RBAM improves kernel-based association testing and disease classification across a broad spectrum of traits by reconstructing disease genotypes and incorporating encoder-decoder weights [6], [7]. as variant-specific representations. Our approach was benchmarked on more than 112,000 real GWAS samples from the UK Biobank, DbGAP, and WTCCC, encompassing 17 diverse diseases and traits, including neurological (Alzheimer's disease, Autism Spectrum Disorder, Schizophrenia), immunological, oncological, cardiometabolic, and quantitative phenotypes. At the core of RBAM is a VAE architecture that projects high-dimensional genotype data into a lower-dimensional latent space, capturing both additive and non-linear relationships. We then use the learned encoder and decoder weights to construct a weighted genetic similarity matrix—applied within a kernel-based association framework to compute a variance-component Q statistic. Empirical results indicate that RBAM-ED (which integrates encoder + decoder weights) outperforms both classical and deep learning baselines, including shapely XAI [8] values from autoencoders, SKAT [9], [10], and REGENIE [11], in terms of power and precision for disease-gene discovery. Furthermore, RBAM methods outperform other methods in the precision of identifying disease-relevant genes in the DisGeNET database. Simulation studies with multiple replicates demonstrate that RBAM maintains a properly calibrated Type I error rate at alpha = 0.05. Beyond association testing, the learned latent representations were fed into machine learning classifiers [12] (including neural networks, logistic regression, random forest, and XGBoost), consistently improving predictive accuracy and AUC over standard polygenic risk scores [13] in 13 complex diseases. To elucidate the biological underpinnings, we performed KEGG pathway and Gene Ontology enrichment, revealing critical processes in immune function, neurodevelopment, and lipid metabolism. Shared gene analyses highlight overlapping pathways in related conditions, such as antigen processing in autoimmune disorders and cholesterol metabolism in Alzheimer's disease and LDL traits. Altogether, RBAM offers a robust framework that leverages unsupervised representation learning to improve association mapping and disease risk prediction in complex traits and diseases.

[1] W. S. Bush and J. H. Moore, "Chapter 11: Genome-Wide Association Studies," *PLoS Computational Biology*, vol. 8, no. 12, p. e1002822, Dec. 2012, doi: 10.1371/journal.pcbi.1002822.

[2] D. J. Schaid, W. Chen, and N. B. Larson, “From genome-wide associations to candidate causal variants by statistical fine-mapping,” Aug. 01, 2018, *Nature Publishing Group*. doi: 10.1038/s41576-018-0016-z.
[3] G. Gibson, “Rare and common variants: twenty arguments,” *Nat Rev Genet*, vol. 13, no. 2, pp. 135–145, Feb. 2012, doi: 10.1038/nrg3118.
[4] G. de los Campos, D. Gianola, and D. B. Allison, “Predicting genetic predisposition in humans: the promise of whole-genome markers,” *Nat Rev Genet*, vol. 11, no. 12, pp. 880–886, Dec. 2010, doi: 10.1038/nrg2898.
[5] Y. Bengio, A. Courville, and P. Vincent, “Representation Learning: A Review and New Perspectives,” Apr. 23, 2014, *arXiv*: arXiv:1206.5538. doi: 10.48550/arXiv.1206.5538.
[6] T. Unterthiner, D. Keysers, S. Gelly, O. Bousquet, and I. Tolstikhin, “Predicting Neural Network Accuracy from Weights,” Apr. 09, 2021, *arXiv*: arXiv:2002.11448. doi: 10.48550/arXiv.2002.11448.
[7] V. Herrmann, F. Faccio, and J. Schmidhuber, “Learning Useful Representations of Recurrent Neural Network Weight Matrices,” Jun. 18, 2024, *arXiv*: arXiv:2403.11998. doi: 10.48550/arXiv.2403.11998.
[8] S. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” Nov. 25, 2017, *arXiv*: arXiv:1705.07874. doi: 10.48550/arXiv.1705.07874.
[9] 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.
[10] M. C. Wu, S. Lee, T. Cai, Y. Li, M. Boehnke, and X. Lin, “Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test,” *The American Journal of Human Genetics*, vol. 89, no. 1, pp. 82–93, Jul. 2011, doi: 10.1016/j.ajhg.2011.05.029.
[11] J. Mbatchou *et al.*, “Computationally efficient whole-genome regression for quantitative and binary traits,” *Nat Genet*, vol. 53, no. 7, pp. 1097–1103, Jul. 2021, doi: 10.1038/s41588-021-00870-7.
[12] D. O. Enoma, J. Bishung, T. Abiodun, O. Ogunlana, and V. C. Osamor, “Machine learning approaches to genome-wide association studies,” *Journal of King Saud University - Science*, vol. 34, no. 4, p. 101847, Jun. 2022, doi: 10.1016/j.jksus.2022.101847.
[13] S. W. Choi, T. S.-H. Mak, and P. F. O’Reilly, “Tutorial: a guide to performing polygenic risk score analyses,” *Nature Protocols 2020 15:9*, vol. 15, no. 9, pp. 2759–2772, Jul. 2020, doi: 10.1038/s41596-020-0353-1.

Quantile-Gated Variational Autoencoder: Applications to High-Dimensional Genomic Data of Small Sample Sizes

Ariel Ghislain Kemogne Kamdoum[1], Janith Weeraman[2], Sasha Chernenkoff[3], Caifeng Li[4], Thierry Chekouo[5], Quan Long[6]

[1] University of Calgary, Alberta, Canada
[2] University of Calgary, Alberta, Canada
[3] University of Calgary, Alberta, Canada
[4] University of Calgary, Alberta, Canada
[5] University of Minnesota, Minneapolis, USA
[6] University of Calgary, Alberta, Canada

Presenting author: arielghislain.kemogn@ucalgary.ca

Keywords: Variational Autoencoder, Quantile-Gated Network, High-Dimensional Data, Small Sample Sizes, Representation Learning, Disease Classification, Genotype-Phenotype Association Study

Variational Autoencoders (VAEs) are powerful tools for capturing hidden representations of complex data through latent distributions in a bottleneck layer. Since their inception, VAEs have relied on a single sample drawn from each latent distribution during both training and inference, a practice grounded in the assumption that the mini-batch size is sufficiently large. However, in applications of VAEs to fields such as genomics, where data are often high-dimensional but of low sample size, limitations arise due to inherently small mini-batches and specific challenges in biological and medical contexts. To address these issues, we propose the Quantile-Gated Variational Autoencoder (qgVAE), which introduces a novel quantile-gated branching network architecture to support multiple samples drawn from each latent distribution, assigning distinct roles to each sample and maximizing the information captured by the latent space. Through mathematical derivations, we demonstrate that qgVAE preserves the variational properties of standard VAEs while mitigating the high variance associated with small mini-batches. Empirical results using real genomic data from seven diseases further validate qgVAE's effectiveness in achieving superior reconstruction accuracy and improved performance in downstream applications.

1. https://www.nature.com/articles/nrg.2016.86
2. https://www.nature.com/articles/nature08800
3. https://www.nature.com/articles/s41592-021-01252-x
4. https://www.nature.com/articles/nrg1916
5. https://www.nature.com/articles/s41587-022-01310-2
6. https://www.nature.com/articles/s41592-023-01787-4
7. https://www.nature.com/articles/s41596-020-0353-1
8. https://www.nature.com/articles/s41588-024-01548-0
9. https://pubmed.ncbi.nlm.nih.gov/15761122/
10. https://doi.org/10.1371/journal.pcbi.1011476
11. https://www.nature.com/articles/s41467-023-42589-2
12. https://arxiv.org/abs/1705.07874
13. https://www.nature.com/articles/s41467-023-36688-y
14. https://www.nature.com/articles/nrg.2016.27
15. https://www.nature.com/articles/s41467-022-29430-1
16. https://www.nature.com/articles/s43586-021-00056-9
17. https://www.nature.com/articles/s41586-021-03243-6
18. https://doi.org/10.1038/s41588-023-01408-9
19. arXiv preprint: 1606.05908

20. arXiv preprint arXiv:1509.00519 (2015)

Mutant Warfare: Cracking the Code of Anti-Phage Resistance in avian pathogenic *Escherichia coli*

Mawra Gohar[1], Matthew Waldner[2], John Fairbrother[3], Matthew Walker[4], Trevor Alexander[2], Charles Martin Nyachoti[5], M. Faizal Abdul Careem[1], Yan D. Niu[1]

[1]Faculty of Veterinary Medicine, University of Calgary, Calgary, Canada
[2]Agriculture and Agri-Food Canada, Lethbridge Research and Development Centre, Lethbridge, Canada
[3]Department of Pathology and Microbiology, Université de Montréal, Montreal, Canada
[4]Canadian Science Centre for Human and Animal Health, Public Health Agency of Canada, Winnipeg
[5]Faculty of Agricultural and Food Sciences, University of Manitoba, Manitoba, Canada

Presenting author: mawra.gohar@ucalgary.ca

Keywords: Avian Pathogenic *Escherichia coli* (APEC), bacteriophage therapy, phage resistance, SNPs, gene loss, defense systems, CRISPR-Cas, phage cocktail

Avian Pathogenic *Escherichia coli* (APEC) causes colibacillosis in poultry, leading to economic losses. With rising antibiotic resistance and reduced antibiotic use, bacteriophage therapy presents a targeted alternative. Our previous study showed that a phage cocktail—vB_EcoM_ASO78A, vB_EcoS_AVIO78A, and vB_EcoS_ASO78B—effectively lysed APEC O78 in broth culture [1] and prevented systemic infection in layers via drinking water. However, intramuscular (IM) administration failed to protect, with phage-resistant mutants emerging in IM-treated birds. In addition, whether or not bacteria develop same anti-phage mechanism between broth culture and in-vivo remains exclusive. We hypothesize that bacteriophage exposure under different biological conditions may select distinct phage-resistant mutants in APEC. Therefore, we aimed to characterize gene mutations and antiviral defense systems (DS) of phage-resistant APEC mutants recovered from three experimental settings. Phage-resistant APEC mutants were recovered from three experimental settings: IM-treated birds, broth cultures exposed to phages (phage killing kinetic assay (PKK), and repeated phage exposures (time kill assay (TKA). Microplate assays confirmed resistance, and fitness analysis showed that resistance to the phage cocktail was weaker ($p = 0.073$) than to single phages. Single-phage resistance, particularly targeting LPS core genes, imposed a higher fitness cost ($p < 0.001$). The whole genome sequencing of 52 resistant mutants revealed single nucleotide polymorphism (SNPs) in 33 isolates, with no core genome mutations in vivo. PKK and TKA mutants exhibited SNPs in 10 and 14 core genes, respectively, with 4 genes overlapping (*OmpA, manC1_3, YfaL,* and *algC_2*). Gene loss analysis identified 40 absent genes across mutants, gene loss was observed in 5 accessory genes from 4 mutants of the in-vivo assay, 19 genes from 13 mutants in PKK, and 22 genes from 21 strains from TKA. Notably, both TKA and PKK assays shared 5 genes (2 genes encoding for unknown function, *grcA, galF,* and *klcA_1*). These included genes (SNPs and gene loss), linked to phage resistance, i.e., those encoding outer membrane proteins, lipopolysaccharide-related proteins, and cell wall components. Interestingly, no genetic changes were detected from mutants treated with the phage cocktail. Pangenome wide association study (PWAS) linked 38 uncharacterized proteins to single-phage resistance and 6 genes to cocktail resistance. Putative defense systems ranged from 14 to 22 per genome, with CRISPR-Cas spacers detected in vivo. In conclusion, prolonged phage exposure in the TKA assay led to more extensive genetic alterations, including some unique defense systems compared to PKK assay. Regardless of biological conditions, phage cocktails led to minimal genetic changes, piggybacking the phage cocktail was an effective approach to minimizing phage resistance.

1. https://www.mdpi.com/1999-4915/15/10/2095

Characterizing the genomics of *Culex pipiens*, including new chromosome-level assemblies of *Cx. p. pipiens* and *Cx p. molestus.*

Leah Jackson[1], Gen Morinaga[1], Andrea Gloria-Soria [2], John Soghigian[1]

[1]University of Calgary, Calgary, Canada
[2]The Connecticut Agricultural Experiment Station, New Haven, CT, USA

Presenting author: leah.jackson1@ucalgary.ca

Keywords: Mosquito, *Culex*, genome assembly, genomics, Hi-C, PacBio sequencing

The West Nile vector, *Culex pipiens*, is a globally invasive mosquito experiencing massive range expansions, closely associated with climate change and anthropogenetic disturbances to the ecosystem. *Culex pipiens* has two morphologically identical bioforms, the above ground northern house mosquito (Cx p. pipiens), and the subterranean London underground mosquito (*Cx. p. molestus*). Culex p. pipiens requires a bloodmeal before egg production, while *Cx. p. molestus* does not. Previous genetic work has shown these two bioforms are genetically distinct and can hybridize, however, little is known about the genetics underlying the subterranean and facultative autogenous attributes of *Cx. p. molestus.* Using PacBio whole genome sequencing and Hi-C scaffolding, this project aims to (1) produce chromosome level genomes of *Cx. p. pipiens* and *Cx. p. molestus*, (2) identify the genomic differences between the two bioforms, and (3) perform comparative genomics with other *Culex* chromosome-level genomes. At this point, we have assembled the 540Mb Culex p. pipiens and the 531Mb *Cx. p. molestus* genomes and are in the process of annotation. Results from this research will not only improve our understanding of mosquito biology and evolution but will add more genomic targets for surveillance and control of *Cx. pipiens*.

4-1BB Expression After Agonist Selection Regulates The TCRβ^{+} CD8αα^{+} IEL Population

Kevin Joannou[1], Dominic Golec[2], Julia May[1], Zalak Khatr[i1], Troy A Baldwin[1]

[1]Department of Medical Microbiology and Immunology, 6-25 Heritage Medical Research Centre, University of Alberta, Edmonton, Alberta T6G 2S2, Canada.
[2]Cell Signaling and Immunity Section, Laboratory of Immune System Biology, NIAID, NIH, Bethesda, MD, USA.

Presenting author: joannou@ualberta.ca

Keywords: Rodent, T Cells, Intraepithelial Lymphocytes, Cell Differentiation, Thymus, scRNA-Seq

The vast majority of developing thymocytes undergo apoptosis due to selective processes that ensure generated T cell receptors (TCRs) can productively engage with peptide-MHC complexes, but not to such a high avidity that T cells are autoreactive. However, some T cell subsets such as TCRβ+ CD8αα+ intraepithelial lymphocytes (IEL) are positively selected by agonist TCR signals strong enough to normally cause clonal deletion. The specific TCR and co-receptor signaling requirements for IEL progenitor (IELp) thymocytes to survive agonist selection remains a subject of ongoing investigation. In this study, we generated and analyzed a single-cell RNA sequencing dataset to investigate the development of IELp within the thymus. We found a complete trajectory of adult-derived Type A IELp development starting from the CD4/CD8 double-positive stage, and found that phenotypically Type B IELp in the adult thymus were transcriptionally consistent with being iNKT cells. Gene expression analysis suggested that IELp are a product of stronger and earlier TCR signaling, similar to conventional CD4 T cells, and those that survive agonist selection receive the strongest TCR signals. We identified 4-1BB as transiently expressed on IELp that receive the strongest TCR signals during agonist selection, and it is expressed again on TCRβ+ CD8αα+ IEL, where it regulates the size of the TCRβ+ CD8αα+ IEL population with pro-survival signals. Thus, some high threshold of TCR signalling results in survival of thymocytes instead of clonal deletion, and correlates with 4-1BB expression that is required for regulating the size of the TCRβ+ CD8αα+ IEL population.

Comparative Genomics of the Lipid Droplet-Associated Protein Seipin Across Eukaryotic Diversity Illuminates an Ancient Origin and Conserved Structural Diversity

Emily M. Kingdon[1,2], Sarah M. Orton[1] , Elisabeth Richardson[1]

[1] Mount Royal University, Calgary, Alberta
[2] University of Alberta, Edmonton, Alberta

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

Lipid droplets (LDs) are an organelle ubiquitous across living organisms (Lundquist et al., 2020). Essential for lipid storage and metabolism, little is known about the evolutionary diversity of LDs, or whether extant LDs share a common origin. It has also been shown in model Metazoa that LDs facilitate an impressive variety of cellular functions, including stress responses, cellular signalling, and membrane remodeling. The distribution of these functions across eukaryotic diversity is unknown. We have examined the evolution of seipin, a protein associated with LD biogenesis, across eukaryotic diversity (Hirst et al., 2014 ; Burki et al., 2020). We have identified a pan-eukaryoticdistribution of seipin, with an evolutionary pattern that indicates presence in the Last Eukaryotic Common Ancestor. Ancient conservation of multiple variants of seipin suggests that seipin may have at least two conserved functional roles within the cell. Finally, we identify a lack of sequence homology between BSCL2/seipin in Homo sapiens and the sei-1 gene previously identified as a functional homologue of BSCL2 in Saccharomyces cerevisiae. Though our results suggest that sei-1 may be a highly divergent homologue of BSCL2 / seipin rather than an unrelated protein, the divergence suggests that researchers may need to be more cautious applying results obtained in Saccharomyces spp. outside of this clade.

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

GWAS of a Dermacentor andersoni Laboratory Colony

Justyna Kruczalak[1], Shaun Dergousoff[2], John Soghigian[1]

[1] University of Calgary, Calgary, Canada
[2] Agriculture and Agri-Food Canada, Lethbridge, Canada

Presenting author: justyna.kruczalak@ucalgary.ca

Keywords: Population genomics, GWAS, tick-paralysis, ticks, Dermacentor andersoni

The Rocky Mountain wood tick (*Dermacentor andersoni*) is a North American tick found predominantly west of the Rocky Mountains to the Pacific coast [1,2]. Tick paralysis is a neurological condition caused by the bite of a Rocky Mountain wood tick, and is characterized by muscle weakness, fatigue, ascending paralysis, and difficulty breathing [3,4]. Within western Canada, past research suggested that two populations of this tick differ in their paralytic ability [5]. *Dermacentor andersoni* found in prairies do not induce paralysis, but ticks found in the mountains typically do [6,7]. It is known that paralytic ability has a genetic basis [8], but the specific gene(s) responsible that control it is still unknown. A genome-wide association study (GWAS) was conducted on *Dermacentor andersoni* from a laboratory colony to investigate statistically significant associations between tick paralysis and regions of the genome. We accounted for potential spurious associations due to population structure. Population genomic statistics were utilized to compare genetic variation between populations that cause paralysis, those that do not, and their hybridized progeny. Our results will be used to assess the impact associated with Dermacentor andersoni and their public health implications in British Columbia.

1. https://doi.org/10.1139/z67-066
2. https://doi.org/10.1093/jmedent/43.1.17
3. https://doi.org/10.1017/S0031182003004670
4. https://doi.org/10.1002/9781118884799.ch11
5. https://doi.org/10.1093/jmedent/22.1.28
6. https://doi.org/10.1603/0022-2585-40.2.185
7. https://doi.org/10.1093/jmedent/45.6.1064
8. https://doi.org/10.1603/me09223

Biogeography of Giant Viruses in the Canadian Arctic Ocean Determined by Metagenomic Sequencing

Yi Lan[1], Srijak Bhatnagar[1,2], Whitney England[1], María Bautista[1,] Casey Hubert[1]

[1]Department of Biological Sciences, University of Calgary, Calgary, Alberta, Canada
[2]Faculty of Science and Technology, Athabasca University, Athabasca, Alberta, Canada

Presenting author: yi.lan@ucalgary.ca

Keywords: Nucleocytoviricota, biogeography, Arctic Ocean, metagenomics, community structure

Giant viruses, belonging to the Nucleocytoviricota phylum, are ubiquitous in marine ecosystems where they play important ecological roles. Previous studies suggest that giant viruses have successfully adapted to cold environments, leading to a flourishing giant virus community in polar oceans. However, their diversity, distribution, and abundance in these high-latitude waters remains largely unknown. In this study, we conducted large-scale sampling in the Canadian Arctic Ocean to investigate the biogeography of giant viruses. Sampling encompassed five different habitats: sea ice brine channels, bulk sea ice, surface seawater, bottom seawater and sediment. Sampling locations spanned multiple oceanic regions including the Labrador Sea, Hudson Bay, Hudson Strait, Baffin Bay and Queen Maud Gulf. From the resulting 115 metagenomic datasets, we successfully assembled 195 metagenome-assembled genomes (MAGs) of giant viruses spanning 5 different taxonomic orders. Among these orders, Imitervirales was most dominant, comprising 81.2% of the total MAGs, followed by Algavirales, Asfuvirales, Pandoravirales and Pimascovirales. Beta diversity analyses revealed significant differences in community composition across habitats with the exception of bulk sea ice and brine channels that exhibited similar viral communities. Alpha diversity analyses indicated that bottom water has the highest species richness of giant viruses, particularly in the Labrador Sea. Sediments harbored the lowest diversity. Furthermore, bathymetric water depth emerged as the most influential factor shaping the giant virus community structure of different habitats in the polar ocean, especially at Labrador Sea. Our study provides valuable insights into the biogeography of giant viruses in the Arctic Ocean.

A Comprehensive Bioinformatics Pipeline for Bacterial Genome Analysis: Multi-Platform Sequencing, Virulence and AMR Profiling for Public Health Applications

Sylvia Li[1,2], Lingling Jin[1], Mohamed Helmy[1,2,3,4]

[1] Department of Computer Science, University of Saskatchewan, Saskatoon, SK, Canada
[2] Vaccine and Infectious Diseases Organization (VIDO), University of Saskatchewan, Saskatoon, SK, Canada
[3] School of Public Health, University of Saskatchewan, Saskatoon, SK, Canada
[4] Department of Computer Science, Lakehead University, Thunder Bay, ON, Canada

Presenting author: mohamed.helmy@usask.ca

Keywords: Bacterial Genomics, Bioinformatics Pipeline, NGS, Antimicrobial Resistance, Taxonomic Classification, Genome Assembly

Advancements in next-generation sequencing (NGS) technologies have transformed bacterial genomics, enabling large-scale genomic studies to investigate pathogenicity, antibiotic resistance, and genetic diversity. To streamline and enhance bacterial genome analysis in public health research, we are developing a comprehensive bioinformatics pipeline capable of handling multi-platform sequencing data, including Nanopore, Illumina, and PacBio reads. The pipeline begins with a quality assessment and control module, utilizing FastQC, SeqKit, and NanoPlot to evaluate raw read quality. Filtering and trimming steps ensure that only high-quality sequences are retained. The processed reads are then assembled using the assembly module which uses platform-specific tools: Flye for Nanopore and PacBio, and SPAdes, Unicycler, or Megahit for Illumina, ensuring optimal assembly accuracy. To classify the assembled genomes, the pipeline integrates Kraken2 and IDTaxa for taxonomic identification. Then, the assembly quality is validated with CheckM2 and GAMBIT to confirm completeness and accuracy. Functional genome annotation module uses Bakta or Prokka, providing detailed insights into coding sequences, genes, and functional pathways. In addition, the pipeline includes specialized modules for antimicrobial resistance (AMR) detection using AMRFinder, multi-locus sequence typing (MLST), and screening for virulence factors and plasmids with Abricate and MOB-suite. The final stage generates interactive visualizations of the genomic data, offering intuitive representations of bacterial genome structure and key features.
By integrating multi-platform compatibility with robust quality control, taxonomic classification, and functional annotation, our pipeline provides a scalable and efficient solution for bacterial genome analysis. It supports a broad range of applications, including outbreak investigation, antimicrobial resistance surveillance, and bacterial phylogenomics.

Explainable Representation Learning Reveals Critical Genes in Cancers

Caifeng Li[1], Jingjing Wu[1], Quan Long[1,2,3,4,5], Alexander Platt[6], Qingrun Zhang[1,2,3]

[1] Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada
[2] Department of Biochemistry and Molecular Biology, University of Calgary, Calgary, Alberta, Canada
[3] Alberta Children's Hospital Research Institute, University of Calgary, Calgary, Alberta, Canada
[4] Mathison Centre for Mental Health Research and Education, Hotchkiss Brain Institute, University of Calgary, Calgary, Alberta, Canada
[5] Department of Medical Genetics, University of Calgary, Calgary, Alberta, Canada
[6] Department of Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, United States of America

Presenting author: caifeng.li@ucalgary.ca

Keywords: Representation Learning, Critical Genes, Tajima's D, eXplainable Artificial Intelligence, hub genes

Representation Learning (RL) is pivotal in high-dimensional data analysis. Previously eXplainable Artificial Intelligence (XAI) has been used to link the RL-learned latent space back to input features to improve the interpretability of black-boxes. In the reverse direction, our group has recently pioneered the use of "interpretability" to quantify the importance of input features. Using RNA data, we showed that genes contributing substantially to latent spaces are functionally relevant, with higher enrichment in disease databases than standard differentially expressed genes and hub genes do. In this work, we extend the model to evolutionary analysis, showing that genes prioritized by "interpretability" are under historical selection related to cancers, quantified by neutrality tests including Tajima's D. These works jointly reveal the novel role of "interpretability", which quantifies feature importance in complex models, similar to the variance component in standard linear analysis.

Early Detection of Cognitive Strain and Mental Fatigue Using Webcam-Based Visual Analysis

Mojtaba Kanani[1], Qingrun Zhang[1]

[1] University of Calgary, Alberta, Canada

Presenting author: mojtaba.kananisarche@ucalgary.ca

Keywords: Facial Video Analysis , Visual Markers, Preventive Monitoring, Deep Learning, Neurophysiological Correlation, EEG, Cognitive Strain , Cross-Modal Learning

This research project aims to develop a non-invasive system for detecting cognitive strain and mental fatigue in individuals working with computers, with particular attention to child users. Using only webcam video input, the proposed framework will identify visual markers associated with elevated levels of cognitive load and mental exhaustion. Unlike existing approaches that rely on predefined behavioral indicators or self-reporting, this research will establish ground truth through correlation with neurophysiological measurements, creating a more objective foundation for detection.

The methodology involves analyzing multimodal datasets containing both facial videos and corresponding brain activity measurements to identify reliable visual indicators of cognitive strain. These indicators will then be used to train deep learning models capable of real-time assessment using only video input. The resulting system could serve as an early warning mechanism to prevent excessive mental strain, potentially reducing the risk of stress-related cognitive issues, particularly in developing brains.

While our current focus is cognitive strain detection, this research framework has broader applications. Since various cognitive states such as anxiety, cognitive overload, physical tiredness, burnout, and limited alertness all produce distinctive patterns in brain activity, the proposed pipeline could be extended to detect these conditions as well. This versatility increases the potential impact of the system across multiple domains where cognitive state monitoring is valuable.

Unlike existing methods that detect fatigue only at critical moments (such as drowsiness detection systems that alert drivers after eye closure is already occurring), our approach identifies earlier indicators of cognitive strain. This preventive stance allows for intervention before users reach dangerous levels of fatigue, creating fundamentally safer systems. For instance, rather than alerting a driver who is already dangerously drowsy, our system aims to detect the earlier signs of fatigue, enabling timely rest breaks before safety is compromised.

Utilizing Image-Expression Directed Linear Mixed Model to discover the genetic variants associated with Brain Disorders

Kahanda Liyanage[1], Rushani N K P[1], Wang. Dinghao[1], Weeraman. Janith[1], Zhang. Qingrun[1]

[1] University of Calgary, Calgary, Alberta

Presenting author: rushaninilakshikakum@ucalgary.ca

Keywords: Linear Mixed Model, Bayesian Sparse Linear Mixed Model, Multi omics, IEDLMM, Image expression, Brain disorders, Bipolar Disorder, Major Depression Disorder, Schizophrenia, Autism Spectrum Disorder

Identifying genetic variants statistically associated with specific diseases is the focus of Genome-Wide Association Studies (GWAS). Advancements in omics technologies have enabled the use of multi-omics data to bridge the gap between genotypes and their resulting phenotypes. Recently, various models have been proposed to utilize omics data for estimating polygenic terms. For example, the Image Mediated Association Study (IMAS)1 leverages brain imaging data to conduct association mapping in legacy GWAS cohorts. Meanwhile, the Expression-Directed Linear Mixed Model (EDLMM)2 incorporates expression data to identify low-effect genetic variants, demonstrating superior performance in terms of power and real data analysis outcomes. However, most current association studies focus on a single biological unit. In our work, we developed an Image Expression Directed Linear Mixed Model (IEDLMM) which utilizes informative weights learned from training genetically predictive models for brain images using a linear mixed model and for gene expressions using a Bayesian Sparse Linear Mixed Model, to estimate the polygenic term in a linear mixed model. Through Simulations we have proven that IEDLMM exhibits higher power than current methods while keeping the type-I error rates under control. By leveraging the UK Biobank image derived phenotypes (IDPs) and GTEx gene expression data, the IEDLMM identified 15 unique genes related to brain disorders across four datasets which are validated through DisGeNET functional annotations proving the efficacy of IEDLMM compared to existing methods. The creation of IEDLMM paves the way for additional exploration in the integration of multiple omics data within a single framework. This method not only improves the credibility of the results but also furthers our knowledge in the field, laying a foundation for future research efforts.

1. https://pubmed.ncbi.nlm.nih.gov/38118447/

2. https://pubmed.ncbi.nlm.nih.gov/38314848/

WISEFLU: An Interactive Platform for Automated Avian Influenza Surveillance and Visualization

Mahafujul Hamid Ananda[1], Lingling Jin[1]*, Mohamed Helmy Shehata[1],2*, Mathew Rogers[2]

[1] Department of Computer Science, University of Saskatchewan, Saskatoon, Canada
[2] Vaccine and Infectious Disease Organization (VIDO), Saskatoon, Canada
* Corresponding author

Presenting author: inquiry@mhananda.info | sez876@usask.ca

Keywords: Avian influenza, Genomic surveillance, Data visualization, Bioinformatics, Phylogenetics, Pandemic preparedness, Automation, Interactive visualization

Avian influenza viruses pose substantial threats to global public health and agricultural systems due to their rapid evolution and zoonotic potential. Effective surveillance requires sophisticated computational methods, yet existing platforms often lack integration between data collection and analysis pipelines or sufficient focus on avian-specific monitoring. We present wiseFlu, a novel web-based platform at https://wiseflu.usask.ca/ that automates the entire surveillance workflow from data acquisition to analysis and visualization, addressing critical gaps in current avian influenza monitoring systems. wiseFlu implements an end-to-end pipeline that begins with automated retrieval of viral sequences and metadata from GISAID using PyAutoGUI automation. The platform performs comprehensive data standardization, including date normalization, geographic information extraction with coordinate mapping, and host taxonomy classification. Processed data is stored in a MongoDB database optimized for efficient querying across multiple dimensions. For phylogenetic analysis, wiseFlu integrates with Nextstrain's augur pipeline, implementing customized preprocessing steps to enhance tree construction quality. The frontend, developed using the MERN stack with D3.js visualizations, provides researchers with a suite of interactive tools including temporal trend analysis, geographic distribution mapping, and phylogenetic exploration through a unified interface. The system's architecture emphasizes modularity and extensibility, allowing for continuous enhancement of analytical capabilities. Key innovations include: (1) fully automated data acquisition from repositories with limited API access; (2) sophisticated data cleaning protocols specifically designed for avian influenza metadata; (3) seamless integration of sequence analysis with contextual metadata; and (4) coordinated visualization components that maintain context across different analytical perspectives. Performance evaluation confirms the platform's capacity to efficiently process and visualize thousands of avian influenza sequences while maintaining responsiveness. wiseFlu represents a significant advancement in avian influenza surveillance by bridging the gap between data collection and actionable insights through automation and interactive visualization. By providing an integrated environment for exploring temporal patterns, geographic spread, and evolutionary relationships simultaneously, the platform enhances researchers' ability to monitor viral evolution and detect potential threats. Future development will focus on implementing advanced mutation analysis capabilities and predictive analytics to further support pandemic preparedness efforts.

1. Shu, Y., & McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data – from vision to reality. Euro Surveillance, 22(13), 30494. https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494
2. Hadfield, J., et al. (2018). Nextstrain: real-time tracking of pathogen evolution. Bioinformatics, 34(23), 4121-4123. https://doi.org/10.1093/bioinformatics/bty407
3. Katoh, K., & Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution, 30(4), 772-780.https://doi.org/10.1093/molbev/mst010

Differential adenosine to inosine RNA editing of SINE B2 non-coding RNAs in mouse unveils a novel type of epi-transcriptome response to amyloid beta neuro-toxicity

Liam Mitchell[1,2], Luke Saville[2], Travis Haight[12], Riya Roy[12], Igor Kovalchuk[3], Athanasios Zovoilis[12]

[1] Max Rady College of Medicine, University of Manitoba, Winnipeg, Manitoba, Canada
[2] Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, Manitoba, Canada
[3] Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, Alberta, Canada

Presenting author: liam.mitchell@umanitoba.ca

Keywords: RNA modification, epi-transcriptomics, inosine, repetitive elements, bioinformatics, nanopore

Alzheimer's disease (AD) is neurodegenerative disorder, whose underlying molecular pathology is largely unclear. The study of standard-dogma molecular biology (DNA→RNA→Protein) has largely been unable to resolve the underlying mechanisms of such complex multi-factor diseases. Recent advancements in sequencing technology have made it possible to study other kinds of molecular control mechanisms, such as RNA editing. The type of RNA editing involving conversion from adenosine to inosine (A-to-I) has been connected with cellular function. Specifically in domains of cell-mediated immunity, dsRNA homeostasis, and cellular stress. Among the RNAs that have been reported to harbor a large percentage of A-to-I edits are non-coding RNAs generated from Short Interspersed Nuclear Elements (SINEs), such as B2 RNAs in mouse and Alu RNAs in human. We have recently shown that B2 RNAs can act as riboswitches, regulating gene expression though self-cleavage, and they are abnormally processed during amyloid beta pathology in hippocampal cells, contributing to the transcriptome de-regulation observed in this condition. Here, we present a novel A-to-I editing analysis approach, which is customized for repetitive elements. We demonstrate this method is able to detect changes in A-to-I editing in a cell-line model of reduced editing. This model is validated by orthogonal methods of EndoV sequencing, in addition to qPCR expression of the editing enzyme. We also show signal-level differences in nanopore sequencing of the same samples, further indication at the presence of editing. We then show changes in editing in B2 RNA as an early response to amyloid neural toxicity in both the hippocampi of a mouse model of amyloid beta pathology and a hippocampal cell culture model of amyloid beta toxicity. This data suggests that the recently described mode of gene-expression-regulation through B2 RNA processing may be intertwined with RNA editing, and that the cell may be employing RNA editing as protective mechanism to counteract the effect of increased B2 RNA processing during the initial phases of response to amyloid beta toxicity. Our findings unveil RNA editing of SINE RNAs as an additional level of epi-transcriptome response to amyloid beta neuro pathology, with potential implications for the role of RNA editing of SINE RNAs also in human and Alzheimer's disease (AD).

Characterization of *Mannheimia haemolytica* isolates uncovers multidrug resistance, prophages, and unique antiviral defense mechanisms that may hinder lytic phage isolation.

Sidra Moqaddes[1], Mawra Gohar[1], Frank van der Meer[1], Tim A. McAllister[2], Beverly Morrison[3] and Yan D. Niu[1*]

[1]Faculty of Veterinary Medicine, University of Calgary, AB, Canada
[2]Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, Lethbridge, AB, Canada
[3]Diagnostic Services Unit, Faculty of Veterinary Medicine, University of Calgary, AB, Canada

Presenting author: sidra.moqaddes@ucalgary.ca

Keywords: Bovine respiratory diseases, Multidrug resistance, *Mannheimia haemolytica,* Antiviral defense systems, Prophages

Bovine respiratory disease (BRD) is one of the most economically significant diseases in North American feedlot and ranks second in negative impact on the global beef industry. Bacterial BRD is usually triggered by primary viral infection or some other factors and is usually associated with *M. haemolytica, P. multocida, H. somni* and *M. bovis*. Among all, *M. haemolytica* is of prime importance despite extensive use of commercially available vaccines and antibiotics. Previously, we attempted to isolate lytic phage using well-characterized *M. haemolytica* strains as hosts, but to no avail. Here, we characterize the AMR profiles and genetic determinants of *M. haemolytica* strains (n=49), which influence resistance and pathogenicity. Phenotypic analysis included AMR testing using Sensititre plates. For genotypic characterization, DNA samples were sequenced using Illumina DNASeq 7.0. Genome assemblies were prepared by Shovill Faster SPAdes assembly of Illumina reads and annotation was done by Prokka Prokaryotic genome annotation. Antimicrobial resistance gens (ARG), prophages, and antiviral defense systems were analyzed using ABRICATE, PHASTEST and PADLOC+ defense finder tools, respectively. Whole genome sequencing generated draft genomes between 2.55-2.74MB composed of 94-177 contigs at a coverage of 100X. All isolates exhibited AMR and 69% were multi-drug resistant (MDR) with resistance to at least one antibiotic (n=18) tested. Of the antimicrobials evaluated, resistance to sulphadimethoxine, tylosin tartrate, neomycin, tulathromycin, and tilmicosin was observed in 100%, 67.35%, 61.22%, 48.98% and 18.37% of the isolates, respectively. Phenotypic resistance was associated with the presence of sulfonamide (*sul2*), macrolide (*CRP*) and aminoglycoside (*strA, strB, aph (3')-Ia, aph (6)-Id, aph(3'')-Ib*) ARG in 55%-100% isolates. Analysis of antiviral defense mechanisms revealed that *M. haemolytica* isolates contain 33 types of defense systems with *RM type-I, Cas_class-1-subtype-1-C, PDC-S05* and *GAO_19* being most abundant. Prophage analysis revealed a minimum of 6 to maximum of 11 prophages in individual isolates. These prophage integrate themselves into the bacterial genome and may confer resistance of their host to lytic phage infection through superinfection immunity. Results of these data co-align with our previous failed attempts to isolate lytic phages and with previous studies unsuccessful attempts to isolate lytic phages of *M. haemolytica*. This finding led us to anticipate that either lytic phages for *M. haemolytica* are rare or absent in nature, or that this host bacterium employs antiviral defense systems/prophages that inhibit phage infection and subsequent isolation. This is the first study to characterize the defense systems of *M. haemolytica* and provide an explanation for the absence of lytic phage isolation to date.

1. https://doi.org/10.1128/AEM.01359-19
2. https://doi.org/10.1017/S1466252309990193
3. https://doi.org/10.1186/s13567-022-01086-1

4. https://doi.org/10.1111/jam.12185

A Pan-Genomic and Comparative Genomic Interrogation of the Yellow Fever Mosquito (Aedes aegypti)

Gen Morinaga[1], Andrea Gloria-Soria[2,3], Jeffrey R Powell[3], John Soghigian[1]

[1]University of Calgary, Calgary, Alberta, Canada
[2]The Connecticut Agricultural Experiment Station, New Haven, Connecticut, USA
[3]Yale University, New Haven, Connecticut, USA

Presenting Author: gen.morinaga@ucalgary.ca

Keywords: Comparative Genomics, Pan-genome, mosquito, *Aedes aegypti,* Arbovirus vector

The yellow fever mosquito (*Aedes aegypti*) is a major vector of numerous arboviruses and found in the sub-tropical and tropical regions of the world (*1*). This near-cosmopolitan distribution is largely due to anthropogenic activity—originally found in sub-Saharan Africa, it was transported throughout the world via trade and transportation. Today, we delimit this species into two sub-species: the native-to-Africa *Ae. aegypti formosus* (Aaf) and the globally invasive *Ae. aegypti aegypti* (Aaa) (*1*). In addition to their geographic distributions, these species exhibit distinct ecological and behavioral differences—whereas Aaf are found in rural and sylvatic environments and feed generally on mammals, Aaa are found almost exclusively in urban areas and seek humans for bloodmeals. To understand the genetic basis for these behavioral and ecological differences, we sequenced and assembled the genomes of six taxa: the *Ae. aegypti* sub-species and four other very closely related species. We use these assemblies in comparative and pan-genomic analyses to interrogate how the genomes of these mosquitoes have evolved and how that has affected the phenotypes that they exhibit.

1. 10.1093/biosci/biy119.

Efficient Deep Learning Approach for Video-based Dense Crop Analysis with Minimal Human Annotation

Keyhan Najafian[1], Farhad Maleki[2], Ian Stavness[1*], Lingling Jin[1*]

[1] Department of Computer Science, University of Saskatchewan, Saskatoon, SK, Canada
[2] Department of Computer Science, University of Calgary, Calgary, AB, Canada
[*] Corresponding author

Presenting author: keyhan.najafian@usask.ca

Keywords: Precision Agriculture, Video Object Segmentation, Wheat Head Segmentation, Deep Learning, Semi-Self-Supervised Learning

Deep learning (DL) is a powerful tool for high-throughput plant phenotyping and precision agriculture, demonstrating success in applications ranging from weed detection [1] to precision spraying [2]. Achieving robust and generalizable performance across diverse plants and growth stages demands large-scale, high-quality annotated datasets for supervised DL. Despite advances in high-resolution imaging and aerial data collection, annotation remains a major bottleneck. The intricate nature of densely packed plant structures requires significant time and effort in pixel-accurate annotation, especially for video object segmentation (VOS), where each training sample requires multiple frames. Semi- and self-supervised learning provide efficient alternatives, enabling model training with minimal manual annotation while maintaining robust performance [3]. Among computer vision applications, VOS plays a pivotal role by delivering pixel-accurate object delineation. This capability enables tracking of crop development over time, resolving ambiguities present in single-frame analysis while informing better decisions.

In this study, we present a semi-self-supervised learning methodology and a convolution-based architecture for wheat head VOS. Our model incorporates an attention mechanism to capture both spatial and temporal features, enhancing VOS accuracy. The attention mechanism enhances temporal feature extraction, enabling the model to focus on relevant plant structures across consecutive video sequences. For training data, we prioritized large-scale data synthesis and pseudo-labeling to avoid manual annotation. Specifically, we employed a cut-and-paste synthesis technique, extracting objects from annotated images and compositing them onto real background footage of bare land areas [3]. Our method integrates both frame-level and video-level motion to simulate realistic field conditions, such as plant swaying and environmental motion. For annotation, we used an image-based model for frame-by-frame data annotation, with human experts reviewing the outputs to ensure accurate video selection. To reduce reliance on human annotations, we trained the model using both pixel-accurate synthetic videos and real data with imperfect labels. We adopted a multi-task learning approach, leveraging preceding frames to predict the subsequent frame and its corresponding segmentation mask, which helped stabilize the training process.

We evaluated our models on both manually annotated and pseudo-labeled wheat field test sets. Our top-performing model achieved 79% segmentation accuracy on drone-captured, human-annotated data despite the inherent challenges of agricultural environment. Notably, our approach outperforms state-of-the-art VOS models that rely on precise ground truth masks even during inference. While developed for wheat, our method demonstrates versatility across crop types, highlighting its potential for precision agriculture and high-throughput plant phenotyping.

1. Murad, N. Y., Mahmood, T., Forkan, A. R. M., Morshed, A., Jayaraman, P. P., & Siddiqui, M. S. (2023). Weed detection using deep learning: A systematic literature review. *Sensors*, *23*(7), 3670. https://shorturl.at/zUvmk
2. Khan, S., Tufail, M., Khan, M. T., Khan, Z. A., & Anwar, S. (2021). Deep learning-based identification system of weeds and crops in strawberry and pea fields for a precision agriculture sprayer. *Precision Agriculture*, *22*(6), 1711-1727. https://shorturl.at/QxP2A
3. Najafian, K., Ghanbari, A., Sabet Kish, M., Eramian, M., Shirdel, G. H., Stavness, I., ... & Maleki, F. (2023). Semi-self-supervised learning for semantic segmentation in images with dense patterns. *Plant Phenomics*, *5*, 0025. https://shorturl.at/w7wBn

Machine learning based sex-specific biomarker discovery in Rheumatoid Arthritis.

Igor Kovalchuk[1], Olga Kovalchuk[2], Sweta Rai[3]

University of Lethbridge, Alberta, Canada

Presenting author: sweta.rai@uleth.ca

Keywords: Rheumatoid arthritis, Differential gene expression analysis, machine learning, feature selection, sex-specific biomarkers.

Background
Rheumatoid Arthritis (RA) is a chronic autoimmune disease characterised by synovial inflammation and joint destruction, which progresses over =me and becomes more prevalent in women. Sex differences influence immune responses, with females exhibting stronger type I and type II interferon signalling and humoral responses. While these differences may interact with autoimmune mechanisms, their exact role in RA remains unclear.

Methods
The NCBI GEO database is used in this study to identify sex-specific biomarkers from transcriptome profiles of RA pa=ents and healthy controls using bioinformatics and machine learning. One dataset was utilised for testing and five for training. Normalisation and differential gene expression analysis were conducted separately for males and females (P-value < 0.05) as part of the data preprocessing. Following the identification of important genes by Boruta feature selection, KEGG, Cytoscape, and GO term enrichment were performed. Accuracy, precision, recall, and AUC-ROC were used to assess the performance of machine learning models (Logis=c Regression, Support Vector Machine, Random Forest) that were trained on the data.

Results
The differential gene expression analysis revealed key pathways, cytokine-mediated signalling in males and neutrophil activation in females. Boruta feature selection highlighted 60 significant genes, including three genes, as potential diagnostic biomarkers for RA. The machine learning model was trained using the differential gene expression analysis data. Logis=c regression achieved a training accuracy of 0.86 and a testing accuracy of 0.75 with the ROC value of 0.85 (train) and 0.71 (test). The base Support Vector Machine had a training accuracy of 0.85 and a test accuracy of 0.72 acer hyperparameter tuning; the training accuracy was improved to 0.96, and the tes=ng accuracy was 0.88 with the ROC curve of 0.89 (train) and 0.82 (test). The base Random Forest model had a training accuracy of 1.0 and a testing accuracy of 0.82. Acer hyperparameter tuning, the training accuracy was 0.96, and the tes=ng accuracy was 0.82 with ROC curves of 1.0 (train) and 0.93 (test). Thus, the Support Vector and Random Forest resulted in ROC curves of 0.89(train) and 1.0(train) respec=vely.

Conclusion
Machine learning-driven biomarker discovery enhances RA diagnostics and personalised treatment strategies. Future research will focus on multi-omics data integra=on, validation of findings with larger cohorts, and development of sex-specific AI tools to improve early detection and optimised therapeutic strategies.

1. https://www.ncbi.nlm.nih.gov/
2. https://string-db.org/
3. https://scikit-learn.org/

AI-Assisted Knowledge Curation for Agricultural Bioinformatics Websites

Robyn Woudstra[1], Mark Berjanskii[1], Scott MacKay[1], Tanvir Sajed[1], Sukanta Saha[1], Eponine Oler[1], Ray Kruger[1], David S. Wishart[1]

1 University of Alberta, Edmonton, Canada

Presenting author: rwoudstr@ualberta.ca

Keywords: retrieval-augmented generation, knowledge graph, ontology, large language model, information extraction

The rapid expansion of agricultural bioinformatics literature and data complicates the efficient retrieval of relevant information. We present an integrated framework for AI-assisted knowledge curation, specifically tailored to agricultural bioinformatics websites.

Our approach begins with high-throughput .PDF and .XML parsing, using natural language processing (NLP) techniques. This step extracts entities, relationships, methodological details, and metadata from published articles, technical reports, and white papers. In this way, unstructured text can be reformatted into structured facts. These structured facts are integrated into a knowledge graph, which is used in conjunction with an index of vectorized text, to construct a Retrieval-Augmented Generation (RAG) system. The RAG system provides context-aware and cited responses to user queries, using generative language models to combine retrieved document passages with structured facts from the knowledge graph. To enhance semantic understanding and enable complex question answering, the knowledge graph is built around a domain-specific ontology. The ontology defines entity types of interest (e.g. genes, pathways, environmental stressors, crops), the properties those entities can have, and the relationships and axioms that can exist between entity types. An interactive chatbot interface facilitates conversational exploration of the knowledge base. Users can pose natural-language questions and receive dynamically synthesized, evidence-backed responses with citations to original sources.

Looking forward, we plan to deploy these AI-assisted databases and chatbots to support our ongoing projects: optimizing growth media for cultivated meat, elucidating molecular determinants of plant resilience, and enhancing decision support within the CAAIN Smart Farm Networks. By streamlining knowledge curation and enabling on-demand, interactive retrieval, our system aims to accelerate discovery and innovation across the agricultural biotechnology landscape**.**

Adenosine-to-Inosine editing of SINE RNAs: A Link to Stress Adaptation Pathways

Riya Roy[1,2,3], Travis Haight[1,2,3], Liam Mitchell[1,2,3], Syed Beenish Rufai[1,2], Parisa Rayomand[1,2], Alyona Gerasimova[1,2], Luke Saville[1,2], Igor Kovalchuk[3] , Athanasios Zovoilis [1,2]

Affiliations:
(1) Department of Biochemistry and Medical Genomics, University of Manitoba, Winnipeg, MN, Canada
(2) Cancer Care Manitoba, Winnipeg, MN, Canada
(3) University of Lethbridge, Lethbridge, AB, Canada

Presenting author: riya.roy@uleth.ca

Keywords: Non-coding RNAs, SINEs, RNA modification, Mouse B2, Cellular stress, Sequencing, GLORI

The mammalian genome contains approximately 2% protein-coding regions, raising questions about the function of the remaining 98% non-coding regions. Among these non-coding regions, Short Interspersed Nuclear Elements (SINEs) are one of the most abundant classes of non-coding RNAs, with murine B2 and human Alu elements being the most prevalent[1]. Adenosine-to-Inosine (A-to-I) RNA editing, catalyzed by adenosine deaminase acting on RNA (ADAR) enzymes, is a widespread post-transcriptional modification that primarily targets double-stranded RNA. ADARs predominantly modify Alu and B2 SINEs[3]. Several studies have highlighted the important role of SINE RNAs in cellular responses to stress[2]. Conversely, A-to-I RNA editing has been implicated in the pathogenesis of various neurological diseases. However, the relationship between A-to-I editing and the stability of SINE RNAs remains poorly understood. In this study, we aim to explore the connection between RNA editing and the stability of ADAR's major targets—B2 SINE RNAs—focusing on the correlation between SINE RNA processing and A-to-I editing in response to cellular stress in mice.

References

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.

Real-Time Pathogen Detection with Nanopore Sequencing

Tayab Soomro[1,2*], Samuel Horovatin[2], Tim Dumonceaux[1], Matthew G. Links[2]

[1] Agriculture and Agri-Food Canada, Saskatoon Research & Development Centre, Saskatoon, SK, Canada
[2] University of Saskatchewan, Saskatoon, SK, Canada

[*]Presenting author: tayab.soomro@usask.ca

Keywords: Pathogen detection, Nanopore Sequencing, DNA, Point-of-care, Pathogen alert

Rapid and precise pathogen detection is essential across a wide range of applications including agriculture, wildlife conservation, and clinical diagnostics. Conventional methods like PCR, while effective for specific gene targets, often fail to identify unexpected pathogens. Its effectiveness depends on predefined primers, which may not bind to a target gene if sequence variability alters the primer-binding site, potentially leading to false negatives. Additionally, the presence of the target gene in non-pathogenic organisms can produce false positives. This can result in inaccurate sample diagnosis. Nanopore sequencing, with its capacity to swiftly generate long-read data, provides a robust alternative. However, realizing its full potential requires tools capable of real-time data analysis.

We present nanoCAS (Nanopore Classification and Alerting System), a web-based application integrated with a nanopore DNA sequencer to meet this challenge. nanoCAS enables users to construct a reference sequence database and establish custom alert thresholds for pathogens or sequences of interest. As sequencing data is generated, nanoCAS aligns reads against the database in real-time, triggering alerts to users when predefined thresholds are reached.

A distinguishing feature of nanoCAS is its ability to offer a comprehensive view of a sample's composition by aligning sequencing reads against a custom reference database in real-time, allowing for the detection of multiple pathogens or sequences of interest simultaneously., extending beyond just the targeted sequences. For instance, in agricultural applications, nanoCAS has been successfully used to identify bacterial pathogens such as *Xylella fastidiosa* in grapevines, a pathogen that causes Pierce's disease and has led to significant crop losses in Europe and the United States, and *Brucella* spp. in cattle and bison, which causes brucellosis, resulting in reproductive issues such as abortions and infertility, and threatens the viability of wild bison herds in Canada. Additionally, it is currently being tested for the detection of *Sclerotinia* in canola samples. This capability to detect a spectrum of bacterial and fungal pathogens across plant and animal hosts underscores its versatility and superiority over traditional PCR methods, which are restricted to detecting predefined target sequences.

Performance evaluations affirm nanoCAS's reliability. In environmental DNA samples, it demonstrates the ability to detect pathogens from very little sequencing data. These preliminary findings highlight its precision and adaptability, suggesting its potential as a dependable tool for pathogen detection. With applications spanning clinical diagnostics to agricultural monitoring, it promises to strengthen pathogen detection strategies, establishing itself as an invaluable resource for researchers and practitioners.

Cities to Microbial Communities: Urbanization Shapes the Plastisphere's Microbial Ecology

Jonas M. Stadfeld[1], Sneha Suresh[1,2], Kira L. Goff[2], Srijak Bhatnagar[1,2]

[1] University of Calgary, Calgary, Canada
[2] Athabasca University, Athabasca, Canada

Presenting author: Jonas.stadfeld1@ucalgary.ca

Keywords: Microplastics, Plastic Pollution, Plastisphere, Biodegradation, Freshwater microbiome, Microbial Communities, Microbial Ecology, Pathogens,

Microplastics (MPs), defined as plastics smaller than 5mm, are an emerging and pervasive pollutant whose impacts on human and ecosystem health are the subject of ongoing research area. One understudied aspect of MPs is the community of bacteria present on the surface of microplastics, known as the plastisphere [1]. The plastisphere contains bacteria with a higher abundance of genes associated with pathogenicity [1], antimicrobial resistance [2], and plastic degradation [3]. Hence, the plastisphere is crucial to our knowledge of MPs' impact on human and ecosystem health. The plastisphere remains understudied in freshwater ecosystems, with little known about the ecological factors that shape its microbial community. Urban areas are the largest contributors of MPs to the freshwater environment [4] and are known to increase pathogenicity and antimicrobial resistance in the freshwater microbiome [5,6]. Hence, we studied the plastisphere of the Bow and Elbow Rivers water and sediments around Calgary using 16s rRNA gene sequencing. We found that MPs harbored a statistically distinct and more diverse microbial community than river water alone. Meanwhile, the *in situ* sediment shared ~80% taxa with the plastisphere, suggesting a relationship between the sediment microbiome and the plastisphere. The plastisphere community differences were more pronounced up and downstream from Calgary than between the two rivers upstream from the confluence in Calgary. This difference was driven by differential abundance in many taxa, including Rhodobacteracaea, Comamonadacaea, and Flavobacteriaceae. Network analysis indicated different co-occurrence patterns in upstream and downstream sites, indicating changes in the resilience and complexity of the community. Furthermore, we observed an increased prevalence of potentially plastic-degrading bacteria upstream of the city. Thus, these changes in the prevalence of microbial taxa and network co-occurrence in the downstream plastisphere indicate the effect of the city in shaping the freshwater plastisphere. The influence of the city could be due to a combination of different factors, stormwater drains, treated wastewater, runoff, etc. and hence, future studies should attempt to disentangle these drivers.

GNAT – An Interactive Web Tool for Gene Neighbourhood Analysis

Nolen Timmerman[1], Chantel Trost[2], Lingling Jin[1*]

[1] Department of Computer Science, University of Saskatchewan, Saskatoon, Canada
[2] The Hospital for Sick Children (SickKids), Toronto, Canada
[*] Corresponding author

Presenting author: njt694@usask.ca

Keywords: Operon, Genomic Neighbourhood, Visualization, Homology, Anti-Defence, Immunity

In prokaryotes and viruses, genes are often found in clusters called operons. These clusters are controlled by a single promoter, are transcribed together and then translated into discrete protein products. Genes within individual operons, as well as adjacent operons, are often functionally related; thus, studying the genes within and around an operon could allow biologists to make functional inferences about poorly characterized or unknown genes based on their proximity to functionally annotated genes. The position of a single gene, or an operon, can be described in terms of its genomic neighbourhood (GN), which consists of the gene or operon of interest and the genes upstream and downstream of it. GN analyses have been successfully applied to determine the function of gene products using the guilt-by-association approach. [1] These analyses have also recently been used to facilitate the discovery of over 100 prokaryotic immune systems and revealed links between prokaryotic and eukaryotic immune systems. [2]
Due to the insights that GN analyses can provide, a tool for biologists to easily search, visualize, annotate, and interact with GNs would prove invaluable. Current tools used for visualizing and comparing multiple GNs, such as clinker & clustermap.js [3], AnnoView [4], and LoVis4u [5], require users to manually process and integrate data from homology searches, making it cumbersome to generate interactive, annotatable GN visualizations at scale in a user-friendly manner. In this research, we developed GNAT (Gene Neighbourhood Analysis Tool), an interactive web-based platform available at https://gnat.usask.ca, which streamlines the entire workflow: it automates homology-based retrieval of GNs from large databases, clusters related sequences into functional groups, and delivers an intuitive, web-based interface for dynamic visualization and annotation—all within a single platform. GNAT was designed to be very easy to use for biologists with no computer science experience, and thus requires only the amino acid sequence of a protein of interest as input, and allows them to select one of several databases to query and extract GNs from. GNAT also offers interactive capabilities, such as automatic labelling of the proteins of interest and their homologs, homology group highlighting and product annotation. It also features a significant file size reduction relative to clinker & clustermap.js without data loss, allowing users to visualize more GNs simultaneously. To further aid biologists in exploring evolutionary relationships, GNAT generates Krona charts [6] for phylogenetic distribution visualizations of the organisms from which the GNs were derived.

1. https://doi.org/10.1101/gr.10.8.1074
1. https://doi.org/10.1038/s41579-023-00934-x
2. https://doi.org/10.1093/bioinformatics/btab007
3. https://doi.org/10.1093/bib/bbae229
4. https://doi.org/10.1093/nargab/lqaf009
5. https://doi.org/10.1186/1471-2105-12-385

Copy Number Variant Detection using the Burrows-Wheeler Transform

C. Malcolm Todd[1], Ian McQuillan[1], and Lingling Jin[1*]

[1] Department of Computer Science, University of Saskatchewan, Saskatoon, Canada
* Corresponding author

Presenting author: malcolm.todd@usask.ca

Keywords: Burrows-Wheeler transform, Copy number variation, DNA variation detection, Maximal exact matches, Longest common prefix.

Copy number variants, or CNVs, represent an essential type of large-scale DNA variation. CNVs can be divided into several types, including insertions, deletions, and duplications, depending on how many copies of a subsequence are present and the samples involved. The Burrows-Wheeler transform (BWT), suffix array, and longest common prefix (LCP) array represent valuable data structures for identtifying Maximal Exact Matches (MEMs) among a collection of DNA sequences. BWTs were used initially to compress large-scale repetitive data [1] and have also been applied in genomic sequence alignment [2]. BWTs rearrange the input text sequence as sorted “circular permutations”, which automatically cluster common subsequences into neighbouring positions within the BWT. This feature makes BWTs well-suited for locating maximal exact matches because we can check nearby entries to see how long their shared prefixes are and whether they form a valid MEM. This work proposes an algorithmic approach to locate clusters of MEM subsequences within the BWT and then identify CNVs’ locations within the input sequences from these clusters using a suffix array. To validate the method, we conducted several experiments using simulated sequences designed to mimic population of samples with varying CNVs. The preliminary results indicate that the method accurately identifies ~85% of the breakpoints around CNV regions. Most errors occur when repeated sequences at a breakpoint introduce ambiguity in the correct index position to report. Furthermore, the method effectively segments the DNA sequences of related samples into smaller subsequences, enabling their relationships to be depicted as a graph. In this graph, nodes represent clusters of identical subsequences, while edges connect clusters to adjacent subsequences from the original input sequences. This representation enables CNV detection within a population by comparing the number of subsequences each input sequence contributes to a cluster. It also determines copy number gains or losses without requiring a reference sequence. The graph representation also retains connections between neighbouring subsequences from the input sequences, which helps locate the deletion and copy number loss breakpoints (the last points where the sequences s,ll matched before DNA content was lost or deleted). Although this method is currently limited to CNV detec,on, the results reveal several promising opportunities to explore sequence characteristics through additional research into subsequence clustering.

[1] https://doi.org/10.1016/j.tcs.2007.07.014

[2] https://doi.org/10.1093/bioinforma,cs/btp698

Assessing Prevalence of Antimicrobial Resistance Genes Across Urban Wetlands in Calgary, Alberta

Kiana A. Vadiat[1], Jeella Z. Acedo[1], Elisabeth Richardson[1]

[1]Mount Royal University, Calgary, Canada

Presenting author: kvadi185@mtroyal.ca

Keywords: Antimicrobial resistance, Antimicrobial resistance genes, Wetlands, eDNA, Environmental health

Antimicrobial resistance (AMR) genes pose a growing global health threat as they can be transferred between bacteria and spread antimicrobial resistance. While research on AMR has largely focused on clinical settings, urban wetlands are increasingly recognized as hotspots for resistance gene accumulation due to their exposure to wastewater, industrial runoff, and other anthropogenic pollutants. Our study determined the prevalence, distribution, and abundance of antimicrobial resistance genes across urban wetlands in Calgary, Alberta. The methods quantified AMR prevalence in the Chan Zuckerberg ID (CZID) platform by comparing the environmental DNA (eDNA) gathered from the urban wetlands to a database of AMR genes, CARD. By identifying drivers of AMR gene distribution, this research will contribute to improved water quality management and public health strategies aimed at mitigating the spread of antimicrobial resistance in urban environments. Its findings could have important implications for human, animal, and environmental health. Based on our results we obtained using multiple detection methods through CARD to ensure all detectable AMR genes were found, we observed that there is relative heterogeneity for species but more homogeneity for drug class across urban wetlands in Calgary. We also detected that there are fewer AMR-containing species detected outside the city, higher abundance of *E. coli* in the northwest potentially associated with large amounts of residential areas, and higher abundance of *S. maltophilia* in the Industrial southeast. This project aligns with my commitment to environmental conservation and will contribute valuable insights into how urbanization shapes microbial diversity and the persistence of resistance genes in wetlands.

1. https://doi.org/10.21203/rs.3.rs-4271356/v1
2. https://doi.org/10.1038/ismej.2014.226
3. https://doi.org/10.1093/nar/gkw1004

A statistical framework integrating somatic mutations, transcriptome, and proteome identifying association between proteins and pathways in cancers.

Dinghao Wang[1], Weijia Cai[2,#], Qingrun Zhang[1,3,4,5,#]

[1] Dept. of Mathematics and Statistics, University of Calgary, Calgary, CA
[2] Dept. of MMICB, SIU School of Medicine, Springfield, United States
[3] Dept. of Biochemistry and Molecular Biology (Adjunct), University of Calgary, Calgary, CA
[4] Hotchkiss Brain Institute, Calgary, CA
[5] Alberta Children's Hospital Research Institute, Calgary, CA
#: corresponding author

Presenting author: dinghao.wang@ucalgary.ca

Keywords: somatic mutation, protein, transcriptome-wide association study, linear mixed model, breast cancer

Transcriptome-wide association study (TWAS) is a powerful analytical approach for identifying significant gene-trait associations by integrating Genome-Wide Association Study (GWAS) and gene expression datasets. In recent years, it has made significant progress. However, with advancements in high-throughput technology, an increasing amount of multi-omics data has become available. This presents an opportunity to extend the TWAS framework beyond gene expression by integrating GWAS with additional multi-omics data, such as somatic mutations and protein expression levels.

In this work, we introduce Somatic-TWAS, a novel model that integrates multi-omics data to identify significant protein-pathway associations. Our approach leverages somatic and germline mutations, gene expression, and protein expression to assess the contribution of pathways to a protein's expression level.

We applied our protocol to breast cancer data from TCGA cohorts, demonstrating its ability to identify causal protein-pathway associations. PAX6 emerged as the protein with the most pathway connections. To further validate this finding, we analyzed its impact on the survival of four breast cancer subtypes (Basal, HER2E, LumA and LumB) and investigated other downstream effects. Our results support PAX6 as a key protein influenced by somatic mutations, revealing its previously unreported subtype-dependent role in breast cancer. This finding highlights the power of our method in unraveling complex, heterogeneous diseases like cancer.

By detecting the contribution of somatic mutations to protein regulation, our approach provides new insights into understanding cancer biology and identifying potential therapeutic targets.

PALL: Personalized Assistant for Local Literature

Janith Weeraman[1], Quan Long[1]

[1]University of Calgary, Calgary, Canada

Presenting author: janith.weeraman@ucalgary.ca

Keywords: Literature Management, Literature Discovery, Research tool, LLM, RAG, Open-source

The explosion of publications in recent decades has made keeping up with the literature a Sisyphean struggle where the amount of research read is far outweighed by the number of articles that never came across a researcher's eyes. This struggle is amplified in multidisciplinary fields such as bioinformatics, where a working knowledge of biology, computer science, and statistics is expected. Existing tools introduced by companies such as OpenAI and Google tend to provide overly generic summaries and are tailored towards researching on single queries and not assisting over the course of long projects.

To address these challenges, we introduce PALL (Personalized Assistant for Local Literature), a novel AI research tool designed to strea

mline literature management and discovery. PALL operates entirely locally, ensuring data privacy and security. PALL offers several innovative features: it runs fully locally, ensuring that no data leaves the user's computer; it is open-source, providing transparency and customizability; and it supports integration with various large language model (LLM) inference providers through a standard API. The tool is built around a user-friendly graphical interface, specifically tailored to be used by non-technical users.

PALL allows researchers to tailor AI interactions to specific research topics using custom filters, such as avoiding preprints or limiting results to studies with certain inclusion criteria. Users can import existing citations from their reference managers, enabling PALL to orient itself based on the user's bibliography. This feature ensures that the tool adapts to the user's research focus, providing relevant and personalized literature recommendations as a project progresses.

One of PALL's standout features is its ability to converse with the user's project bibliography. Researchers can ask questions and receive answers based on their existing literature, or incorporate web searches for a broader perspective.

PALL leverages the knowledge understanding capabilities of LLMs to search and curate personalized feeds of new articles. By analyzing the user's bibliography, PALL can recommend articles that are highly relevant to the user's research, keeping them updated with the latest findings in their field.

While PALL is in its pre-alpha stage and is currently being tested internally, initial feedback has been positive. The next steps involve refining the tool's algorithms, expanding its integration capabilities, and conducting user studies to gather feedback and improve the user experience.

Profiling RNA modifications across a wide range of human/mouse tissues through nanopore direct RNA sequencing and modification basecalling

Elly Wu, Jemaneh Habtewold, Alexey Sizykh, Luke Saville, Liam Mitchell, Travis Haight, Parisa Rayomand, Athanasios Zovoilis

Department of Biochemistry and Medical Genetics, University of Manitoba, Canada
Paul Albrechtsen Research Institute, CancerCare Manitoba, Canada
Children's Hospital Research Institute of Manitoba, Canada

Presenting author: elly.wu@umanitoba.ca

Keywords: RNA modification, nanopore direct RNA sequencing, epitranscriptome; modification basecalling

RNA modifications are chemically modified RNA nucleotides. They usually cause structural changes in RNA molecules, affect their stability and interaction with other molecules, and thus regulate gene expression. Many RNA modifications have been identified, and studies focusing on them have kept expanding in recent years. Direct RNA Sequencing (DRS) by Oxford Nanopore Technology (ONT) evolved to read native RNA transcripts and eliminated potential PCR-derived bias. Combining with computational approaches, many tools have been developed to analyse RNA modifications in DRS data with the ability to study multiple types simultaneously.

In this study, we generated DRS data in a wide range of human and mouse tissues, and performed modification calling with the Dorado basecaller to identify four types of modifications: 6-methyladenine, inosine, pseudouridine and 5-methylcytosine. Preliminary analyses highlight those mostly modified genomic sites and tissue-based clustering. In addition, we present modification patterns across various genomic features, such as transcription start/end sites and exon/intron boundaries. This study will make a large variety of DRS data available for the first time, and we aim to illustrate the modification profiles and eventually compile the information into a modification database.

Systematic Mapping of RNA Modifications Across Tissues and Species

Alex Sizykh[1,2], Jemaneh Habtewold[1,2], Elly Wu[1,2], Sani Zaidi[1,2], Athanasios Zovoilis[1,2]

1. Department of Biochemistry and Medical Genetics, University of Manitoba

2. Paul Albrechtsen Research Institute, CancerCare Manitoba

Presenting author: alex.sizykh@umanitoba.ca

Keywords: RNA modifications, RNA modifications atlas, m6A, epitranscriptomics, Oxford Nanopore Technology, Direct RNA Sequencing

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 (m6A), N4-acetylcytidine (ac4C), pseudouridine (Ψ), and 5-methylcytidine (m5C)—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 m6A enrichment patterns reported previously and further demonstrate that deposition of this modification is highly context-dependent, underscoring the need for continued mechanistic investigation.

Dissecting the regulatory mechanism governing SINE sense and antisense RNA expression in response to cellular stress

Alyona Gerasimova[1,2], Syed Beenish Rufai[1,2], Riya Roy[1,2], Parisa Rayomand[1,2], Alex Sizykh[1,2], Travis Haight[1,2], Athanasios Zovoilis[1,2]

1. Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, MB

2. Paul Albrechtsen Research Institute, CancerCare Manitoba, Winnipeg, MB

Presenting author: gerasim1@myumanitoba.ca

Key Words: Non-coding RNAs, SINE RNA, Alu elements, Heat shock, Sequencing.

Alu elements are primate-specific Short Interspersed Nuclear Elements (SINEs) that comprise over 10% of the human genome and are transcribed into non-coding RNAs, particularly under stress conditions [1,2]. While sense-strand Alu RNAs have been shown to modulate transcription through interactions with RNA polymerase II [3], the function of antisense Alu transcripts remains poorly understood. Sense Alu RNAs are endogenously expressed and undergo stress-induced processing [4], suggesting that they may play coordinated roles in regulating gene expression. However, the potential for regulatory crosstalk between the two strands has not been systematically investigated. This project aims to characterize the interplay between sense and antisense Alu RNAs during cellular stress, with a focus on transcriptional regulation, RNA stability, and post-transcriptional processing. Understanding this dynamic could uncover a previously unrecognized layer of non-coding RNA regulation and provide insights into the role of SINEs in cellular homeostasis and disease.

References:

1. Kramerov DA, Vassetzky NS. SINEs. Wiley Interdiscip Rev RNA. 2011 Nov-Dec;2(6):772-86. doi: 10.1002/wrna.91.

2. Deininger P. Alu elements: know the SINEs. Genome Biol. 2011 Dec 28;12(12):236. doi: 10.1186/gb-2011-12-12-236.

3. Cheng Y, Saville L, Gollen B, Veronesi AA, Mohajerani M, Joseph JT, Zovoilis A. Increased Alu RNA processing in Alzheimer brains is linked to gene expression changes. EMBO Rep. 2021 May 5;22(5):e52255. doi: 10.15252/embr.202052255.

4. 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.

Role of ac4C RNA acetylation in cellular response to stress

Parisa Rayomand[1,2] , Elly Wu[1,2], Syed Beenish Rufai[1,2] , Riya Roy[1,2,3], Jemaneh Habtewold[1,2], Alex Sizykh[1,2], Athanasios Zovoilis [1,2]

(1) Department of Biochemistry and Medical Genomics, University of Manitoba, Winnipeg, MN, Canada
(2) Cancer Care Manitoba, Winnipeg, MN, Canada
(3) University of Lethbridge, Lethbridge, AB, Canada

Presenting author: rayomanp@myumanitoba.ca

Keywords: Ac4C, NAT 10, RNA modification, Cellular stress, Sequencing

N4-acetylcytidine (ac4C) is a conserved RNA modification that was first identified in yeast tRNA in 1966. Initially believed to exist only in tRNA and rRNA, but recently ac4C has been detected in mRNA in yeast and human cells. Its presence from archaea to eukaryotes underscores its evolutionary importance. In human cells, NAT10 is the main enzyme responsible for the acetylation of cytidine to ac4C. Ac4C has been shown to have an impact on several essential cellular processes, including RNA stability, translation efficiency, and stress responses. Furthermore, dysregulation of NAT10 activity is associated with aging, inflammation, and various cancers. Overexpression of NAT10 has been linked to tumor progression, making both NAT10 and ac4C attractive candidates for therapeutic targeting and biomarker development.

Recent studies suggest that RNA modifications, including ac4C, play dynamic roles in how cells respond to stress. However, our understanding of the specific function and regulation of ac4C in this context remains limited. To address this gap, this research aims to explore ac4C RNA acetylation as an epi-transcriptomic change and checking it in cellular stress responses.

The first objective of this study is to identify regions of RNA that are modified by ac4C under normal conditions, using Nanopore sequencing in cell lines. The second objective is to investigate how these acetylation patterns change in response to cellular stress, such as oxidative stress or heat shock. Comparison of ac4C profiles between stressed and unstressed cells may reveal shifts in acetylation patterns associated with stress adaptation or cellular dysfunction.

Overall, this study will enhance our understanding of how ac4C contributes to post-transcriptional regulation of RNA during stress responses and may provide insights into its potential role in disease processes.

PIAS1 Regulates Gene Expression in a Cell Type and Context-Dependent Manner and Correlates with Improved Survival in Oral Squamous Cell Carcinoma

Parisa Ghahremanifard[1], Jinsu An[1], Ayan Chanda[1], Angela Chan[3], Doha Itani[4], Steven C. Nakoneshny[5], T. Wayne Matthews[5,6], Shamir Chandarana[5,6], Robert D. Hart[5,6], Joseph C. Dort[5,6,7], Martin D. Hyrcza[2,5], Shirin Bonni[1], Pinaki Bose[1,2,5,7,#].

[1]Department of Biochemistry and Molecular Biology, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada.
[2]Department of Pathology and Laboratory Medicine, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada.
[3]Functional Tissue Imaging Unit, Arthur JE Child Comprehensive Cancer Centre, Alberta Health Services, Calgary, AB, Canada.
[4]Molecular Diagnostics and Cytogenetics, Saint John Regional Hospital, Horizon Health Network, Saint John, NB, Canada.
[5]Ohlson Research Initiative, Arnie Charbonneau Cancer Institute, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada.
[6]Department of Surgery, Section of Otolaryngology – Head & Neck Surgery, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada.
[7]Department of Oncology, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada.

[#] Corresponding author: Pinaki Bose – pbose@ucalgary.ca

Presenting author: parisa.ghahremanifar@ucalgary.ca

Keywords: PIAS1, oral squamous cell carcinoma, tumor microenvironment, single-cell RNA-seq, stromal biomarker

PIAS1, a SUMO E3 ligase involved in multiple signaling pathways, has been associated with oncogenic or tumor-suppressive functions depending on the context[12]. In our recent study, we demonstrated a tumor-suppressive role for PIAS1 in oral squamous cell carcinoma (OSCC) through suppression of the TGFβ signaling pathway. However, the role of PIAS1 in the tumor microenvironment (TME) has not yet been explored. The TME plays a critical role in tumor progression, drug delivery, immune regulation, and metastasis[3]. To investigate the role of PIAS1 in the OSCC TME, we assessed PIAS1 protein expression in patient-derived tissue microarrays (TMAs) and evaluated its transcriptional impact using single-cell RNA sequencing (scRNA-seq).

Immunohistochemical analysis of TMAs from the Ohlson OSCC cohort revealed that high levels of stromal PIAS1 correlates with significantly improved survival ($p = 0.0027$). To further explore cellular functions affected by PIAS1 within the TME, we analyzed publicly available scRNA-seq data from OSCC patients and normal tissues, identifying nine major cell types in tumors and eight in normal samples. Following data integration, PIAS1 expression was assessed across cell types with matched tumor-normal pairs.

Although a higher proportion of stromal and immune cells in OSCC expressed PIAS1 compared to normal tissue, expression levels per PIAS1-positive cell were significantly reduced in cancer-associated fibroblasts (CAFs), tumor-associated macrophages (TAMs), T cells, and endothelial cells in the TME compared to their normal counterparts (adjusted $p < 0.0001$). Differential gene expression analysis between PIAS1-positive and -negative cells was conducted within fibroblasts, macrophages, and T cells, which are the most

abundant non-epithelial cell types in the TME. Normal macrophages exhibited the strongest PIAS1-associated transcriptional response (75 DEGs), followed by T cells from normal (27 DEGs) and tumor tissues (12 DEGs). CAFs and TAMs showed minimal differential expression, indicating a dampened PIAS1 response in these cells. Across all comparisons, PIAS1-regulated genes were largely non-overlapping between cell types and tissue types, suggesting highly cell type and context-specific activity.

Together, these results suggest PIAS1 as a potential modulator of stromal behavior in OSCC. The association of elevated stromal PIAS1 expression with improved patient outcomes highlights its relevance as a biomarker in OSCC.

1. https://pmc.ncbi.nlm.nih.gov/articles/PMC5426774/
2. https://www.nature.com/articles/onc2015292
3. https://www.nature.com/articles/s41467-023-39762-1

Double-Negative T Cells in HNSCC: Identification of Distinct Regulatory and Cytotoxic Subsets with Prognostic Significance

Jinsu An[1,2], Ilana Lebon[2], Min-su Kim[2,3], Beumkeun Cho[4], ChungGyu Park[4], Jongbok Lee[1,2], and Pinaki Bose[1,2]

[1]Department of Biochemistry and Molecular Biology, University of Calgary, Calgary, AB, Canada.
[2]Arnie Charbonneau Cancer Institute, University of Calgary, Calgary, AB, Canada.
[3]Department of Otolaryngology-Head & Neck Surgery, CHA Bundang Medical Center, CHA University School of Medicine, Gyeonggi-do, South Korea
[4]Department of Microbiology and Immunology, Seoul National University, Seoul, South Korea

Presenting author: jinsu.an@ucalgary.ca

Keywords: Double Negative T cell, DNT, Head and Neck Cancer, HNSCC, Single Cell RNA Sequencing, scRNA-seq, Immuno-oncology, Bioinformatics

Introduction

Head and neck squamous cell carcinoma (HNSCC) ranks as the seventh most common cancer globally and is associated with a poor prognosis[1]. This disease is profoundly immunosuppressive, characterized by excessive proinflammatory cytokine secretion and impaired immune effector cell function[2]. While neoadjuvant immunotherapy has brought new hopes for patients with resectable HNSCC, response rates remain low, ranging from 0–17% for anti-PD-1 monotherapy and 20–35% when combined with antiCTLA-4 therapy[3]. Hence, there is an urgent need to better understand the role of immune cells in HNSCC pathogenesis. Here, we report the first identification of functionally distinct double-negative T cell (DNT; CD3+ CD4– CD8–) subsets in HNSCC, offering new avenues for therapeutic intervention.

Methods and Results

Integration of three publicly available single-cell transcriptomic datasets from primary HPV-negative HNSCC tumors revealed the existence of five unique DNT clusters. Two prominent, functionally diverse subsets were identified: a *FOXP3*+ regulatory DNT subset and a *GZMB*+ cytotoxic DNT subset. In bulk RNAseq data from the HNSCC cohorts within Cancer Genome Atlas (TCGA), high expression of the regulatory-DNT signature correlated with significantly reduced progression-free interval (p=0.034) and disease-specific survival (p=0.043). Conversely, high expression of the cytotoxic-DNT signatures correlated with improved patient outcomes (p=0.0084). Functional assays confirmed these transcriptomic findings: CD25+ FOXP3+ DNTs significantly suppressed CD8+ T cell proliferation, while *ex vivo* bulk-expanded DNTs demonstrated potent cytotoxic activity against CAL33 and UMSCC29 HNSCC cell lines.

Conclusion

Our study unveiled a novel dual functional role for DNTs in HNSCC, with subsets exhibiting either tumor-promoting or tumor-suppressing properties. We continue to further characterize these subsets in HNSCC patient samples, highlighting their potential as prognostic biomarkers and innovative immunotherapeutic targets.

1. https://acsjournals.onlinelibrary.wiley.com/doi/10.3322/caac.21660
2. https://ascopubs.org/doi/10.1200/JCO.2015.61.1509
3. https://www.sciencedirect.com/science/article/pii/S1567576923006525?via%3Dihub