The 2nd Annual BioNet Conference

May 25-27, University of Calgary, Alberta, Canada

The Western Canada Bioinformatics and Omics conference

Talks. Networking. Poster Sessions.

Conference Program and Abstract Book

The conference organizers thank the following sponsors for their kind support of BioNet and the 2nd Annual BioNet Conference:

UNIVERSITY OF CALGARY

University of Lethbridge

UNIVERSITY OF CALGARY
CUMMING SCHOOL OF MEDICINE
Centre for Health Informatics

Southern Alberta Genome Sciences Centre

GenomeAlberta

GenomeCanada

A special thanks to our generous industrial sponsors:

Oxford NANOPORE Technologies

PacBio

Welcome to the 2nd Annual BioNet Conference

Dear members of the BioNet community,

On behalf of the organizing committee, we would like to welcome you to the 2nd Annual BioNet Alberta Bioinformatics and Omics Conference. BioNet Alberta is a diverse network-based approach to building provincial capacity in the area of Bioinformatics and Computational Biology. We seek to achieve these objectives by fostering an environment of collaboration and cooperation among researchers across a wide variety of disciplines. This conference represents an important physical step in bringing together scientists from institutions across Alberta and the Western provinces to identify and promote services, platforms and potential partnerships that may be mutually beneficial. We hope you enjoy this year's conference!

2nd Annual BioNet Conference Organizing Committee:

Eric Merzetti	Angeliki Pantazi	Athanasios Zovoilis
BioNet Alberta	University of Lethbridge	University of Lethbridge

2nd Annual BioNet Conference Scientific Organizing Team:

Jason de Koning	Paul Gordon	Emily Herman	Athanasios Zovoilis
University of Calgary	University of Calgary	University of Alberta	University of Lethbridge

2nd Annual BioNet Conference University of Calgary Conference Service Facilitators:

Jeffrey Rothenbusch Kathy Fergusson

2nd Annual BioNet Conference Student Organizing Committee:

Luke Saville Riya Roy Travis Haight Elly Wu

Contact Information:

Email: bionet@uleth.ca

Twtter: @albertabionet / Conference Hashtag #BioNetAB2022

Website: www.bionet-meeting.org / www.bionet-alberta.org

Conference Venue and Locations:

The 2nd Annual BioNet Conference will be held at the main campus of the University of Calgary in the Science Theatres (798 Campus Pl NW #700, Calgary, AB T2N 4V8) and Taylor Institute (434 Collegiate Blvd NW, Calgary, AB T2N 1N4).

The map below shows the locations of these buildings:

All talks, breakfast, lunches, and poster sessions will take place in the Science Theatre Building (ST140/148).

Dinners will take place in the Taylor Institute (TI160).

Additional information including park maps may be found on our website www.bionet-meeting.org/traveltips

The 2nd Annual BioNet Conference

Wednesday, May 25th

12:00pm-3:30pm	Registration	ST140/148 Hallway
1:00pm-1:25pm	Opening Ceremony	ST140
1:30pm-2:45pm	Session 1	ST140
3:00pm-3:30pm	Refreshment Break	ST140/148 Hallway
3:30pm-4:30pm	Keynote: Dr. Fiona Brinkman	ST140
4:30pm-5:40pm	Session 2	ST140
6:00pm-7:00pm	Dinner	TI160

Thursday, May 26th

8:00am-9:00am	Breakfast	ST140/148 Hallway
9:00am-10:50am	Session 3	ST140
10:55am-11:25am	Refreshment Break	ST140/148 Hallway
11:25am-12:45pm	Session 4	ST140
1:00pm-2:00pm	Lunch	ST140/148 Hallway
2:00pm-3:00pm	Poster Session	ST140/148 Hallway
3:00pm-4:00pm	Session 5	ST140

4:00pm-4:30pm	Refreshment Break	ST140/148 Hallway
4:30pm-5:55pm	Session 6	ST140
6:00pm-7:00pm	Dinner	TI160

Friday, May 27th

8:00am-9:00am	Breakfast	ST140/148 Hallway
9:00am-10:45am	Session 7	ST140
10:55am-11:25am	Refreshment Break	ST140/148 Hallway
11:25am-12:55pm	Session 8	ST140
1:00pm-2:00pm	Lunch	ST140/148 Hallway
2:00pm-3:00pm	Poster Session	ST140/148 Hallway
3:00pm-3:30pm	Refreshment Break / Group Photo	ST140/148 Hallway
3:30pm-4:30pm	PI / Student Meetings	ST140 & ST148
4:30pm-5:30pm	Keynote: Dr. Steven Jones	ST140
5:30pm-7:00pm	Seated Dinner and Awards	TI160
7:30pm-?	Unofficial Post-Conference Gathering	Kilkenny Irish Pub

2nd Annual BioNet Conference – Full Program

Wednesday, May 25th

Registration starting at 12pm / badge and poster number pickup in Hallway outside ST140/148

*Numbers indicate abstract book page number for speaker / poster presenter

Opening Ceremony (ST140)

1:00 – 1:05pm	**Dr. Athanasios Zovoilis** Academic Lead, BioNet Alberta
1:05 – 1:10pm	**Dr. André Buret** Associate Vice President (Research), The University of Calgary
1:10 – 1:15 pm	**Dr. Dena McMartin** Vice President (Research), The University of Lethbridge
1:15 – 1:20 pm	**Dr. Gijs van Rooijen** Chief Scientific Officer, Genome Alberta
1:20 – 1:25 pm	**Dr. Eric Merzetti** Network Manager, BioNet Alberta

Session 1: Spotlight Session (ST140)

Chair: Dr. Eric Merzetti

		*
1:30 – 1:50 pm	**Dr. Quan Long, The University of Calgary** Discovering genetic basis of complex traits by utilizing "borrowed data-bridge"	**9**
1:50 – 2:10 pm	**Rodrigo Ortega-Polo, Agriculture and Agri-Food Canada** The Computational Analyses Research Support Unit of the Lethbridge Research and Development Centre	**10**
2:10 – 2:30 pm	**Dr. Tarah Lynch, Alberta Precision Laboratories** COVID-19: A Catalyst for Change	**9**
2:30 – 2:45 pm	**Short talk: Dr. Ana Nikolic, University of Calgary** Sub-clonal Copy Number Variation and Chromatin State Heterogeneity in Cancer Single-cell Chromatin Accessibility Data	**18**

3:00 – 3:30 pm	Refreshment Break (Hallway Outside ST140/148)	

Opening Keynote Lecture: Dr. Fiona Brinkman (ST140)

Distinguished Professor, Molecular Biology and Biochemistry Dept, Associate Professor, School of Computing Science, and Faculty of Health Sciences Simon Frasier University

Chair: Dr. Paul Gordon

3:30 – 4:30 pm	Pleasing People and Pathogens: Bioinformatics Battles	**1**

Session 2: Omics & Human Disease 1 (ST140)

Chair: Dr. Athanasios Zovoilis

4:30 – 4:50 pm	**Dr. Marco Gallo, University of Calgary** Title TBD	**5**
4:50 – 5:10 pm	**Dr. Olga Kovalchuk, University of Lethbridge** Title TBD	**8**
5:10 – 5:25 pm	**Short talk: Reid McNeil, University of Calgary** Evaluating DNA Amplifications and their effect on Squamous Cell Carcinoma Prognosis using The Cancer Genome Atlas	**16**
5:25 – 5:40 pm	**Short talk: Dr. Michael J Johnston, University of Calgary** Type B Ultra Long-Range Interactions in PFAs (TULIPs) represent recurrent epigenomic alterations in pediatric ependymoma	**42**
6:00 – 7:00 pm	Dinner (TI160)	

Thursday, May 26th

8:00 – 9:00 am	Breakfast (Hallway Outside ST140/148)	

Session 3: Omics & Human Disease 2 (ST140)

Chair: Dr. Laura Sycuro

9:00 – 9:20 am	**Dr. Edwin Wang, University of Calgary** Predictive genomics, from understanding of genomics to predictive models	**12**
9:20 – 9:40 am	**Dr. Laura Sycuro, University of Calgary** Piercing the fog of the gut-brain axis with metagenomics	**11**
9:40 – 10:00 am	**Dr. Pinaki Bose, University of Calgary** Title TBD	**4**

10:00 – 10:20 am	**Dr. Gerlinde Metz, University of Lethbridge** Can't buy me love: epigenetic and metabolomic determinants of health linked to social isolation	**10**
10:20 – 10:35 am	**Short talk: Dr. Ankita Narang, University of Calgary** Whole-genome sequencing analysis of clozapine-induced myocarditis	**40**
10:35 – 10:50 am	**Short talk: Lilit Antonyan, University of Calgary** Genome-Wide Association Study of Pediatric Obsessive-Compulsive Behaviors, the Correlation with Imaging Endophenotypes	**14**
10:55 – 11:25 am	Refreshment Break (Hallway Outside ST140/148)	

Session 4: Microbial Omics (ST140)

Chair: Dr. Ian Lewis

11:25 – 11:45 am	**Dr. Ian Lewis, University of Calgary** Predicting the severity of infections via microbial proteomics: insights from a city-wide survey of blood stream infections over 20 years	**8**
11:45 – 12:00 pm	**Short-talk: Dr. Peipei Zhang, AAFC Lacombe** Large scale genomic analysis of Escherichia coli genomes to characterize the transmissible locus of stress tolerance (tLST) and tLST harboring E. coli	**68**
12:00 – 12:15 pm	**Short talk: Adaobi Ojiakor, University of Alberta** The evolutionary diversification of the Salmonella artAB toxin locus	**17**
12:15 – 12:30 pm	**Short talk: Chenhua Li, University of Calgary** A transcriptomic view of the molecular mechanisms that the lancet liver fluke uses to manipulate its hosts behavior	**36**
12:30 – 12:45 pm	**Short talk: Abhinaya Venkatesan, University of Calgary** Identification of a genetically divergent population of the human soil transmitted helminth Trichuris in Cote d'Ivoire that is unresponsive to albendazole-ivermectin combination treatment	**32**
1:00 – 2:00 pm	Lunch (Hallway Outside ST140/148)	

Poster Session 1 (Hallway Outside ST140/148)

2:00 – 2:30 pm	Even Numbered Posters Present
2:30 – 3:00 pm	Odd Numbered Posters Present

Session 5: Nanopore Sequencing (ST140)

Chair: Dr. Athanasios Zovoilis

3:00 – 3:20 pm	**Dr. Emily Herman, University of Alberta** A Snakemake workflow for detecting Bovine Respiratory Disease pathogens and antimicrobial resistance genes in third generation metagenomic sequencing data	**70**
	Dr. Athanasios Zovoilis, University of Lethbridge NERD-seq a novel approach of Nanopore direct RNA sequencing	**13**
	Dr. Paul Gordon, University of Calgary Is it signal or noise? Dissecting raw direct RNA data from nanopore devices	**5**
4:00 – 4:30 pm	Refreshment Break (Hallway Outside ST140/148)	

Session 6: Advanced Genomic Approaches (ST140)

Chair: Dr. Jason de Koning

4:30 – 4:50 pm	**Paul Stothard – University of Alberta** Bioinformatics for genetic improvement, management, and conservation of livestock	**11**
4:50 – 5:10 pm	**Jason de Koning – University of Calgary** The Population Genetics of Non-Equilibrium Molecular Evolution: A Unified Approach to Understanding Neutral and Non-Neutral Evolutionary Processes	**4**
5:10 – 5:25 pm	**Short talk: Dinghao Wang, University of Calgary** cLD: Rare-variant disequilibrium between genomic regions identifies novel genomic interactions	**28**
5:25 – 5:40 pm	**Short talk: Vladimir Avramovic, University of Calgary** Use of reference population data for resolving variants of unknown significance in hematopoietic genes	**20**
5:40 – 5:55 pm	**Short talk: Deshan Perera, University of Calgary** CATE: A fast and scalable CUDA implementation to conduct highly parallelized evolutionary tests on large scale genomic data	**22**
1:00 – 2:00 pm	Dinner (TI160)	

Friday, May 27th

8:00 – 9:00 am	Breakfast (Hallway Outside ST140/148)

Session 7: Agricultural Omics (ST140)

Chair: Dr. Tim McAllister

9:00 – 9:25 am	**Plenary Talk: Dr. Tim McAllister, AAFC Lethbridge** Genomic approaches to defining the molecular ecology of AMR across a “One-Health" continuum	**3**
9:25 – 9:45 am	**Dr. Igor Kovalchuk, University of Lethbridge** Transgenerational response to heat in Arabidopsis thaliana - genetic and epigenetic components	**7**
9:45 – 10:00 am	**Dr. Devin Holman, AAFC Lacombe** The swine gut microbiome and resistome in conventional and organic production systems	**5**
10:00 – 10:15 am	**Short talk: Dr. Elisabeth Richardson, University of Alberta** Microbiome analysis of tailings reclamation protocols in the Athabasca Oil Sands Region	**37**
10:15 – 10:30 am	**Short talk: Dr. Marta Gerasymchuk, University of Lethbridge** Rejuvenation Effects of Phytocannabinoids and Anti-aging Drugs on Dermal Fibroblasts During Aging	**38**
10:30 – 10:45 am	**Short talk: Sani Zehra Zaidi, University of Lethbridge** Genomic characterization of *Enterococcus hirae* from beef cattle feedlots and associated environmental continuum	**27**
10:55 – 11:25 am	Refreshment Break (Hallway Outside ST140/148)	

Session 8: Omics & Human Disease 3 (ST140)

Chair: Dr. James Wasmuth

11:25 – 11:45 am	**Dr. James Wasmuth, University of Calgary** A hitchhiker’s guide to a parasite’s genome	**12**
11:45 – 12:00 pm	**Dr. Maja Tarailo-Graovac, University of Calgary** Genome sequencing and complex disease mechanisms in rare disease patients	**12**
12:00 – 12:15 pm	**Dr. Gane Ka-Shu Wong, University of Alberta** Drug targets and lead compounds in the era of “free” data for humans and biodiverse taxa	**6**
12:15 – 12:30 pm	**Short talk: Dr. Viraj Muthye, University of Calgary** Identification of mimicry between protein structures and its role in host-parasite interactions	**39**
12:30 – 12:45 pm	**Short talk: Luke Saville, University of Lethbridge** Processing of Alu and B2 SINE RNAs by Hsf1 is dysregulated in neuro-toxicity models and Alzheimer’s disease patients	**29**
1:00 – 2:00 pm	Lunch (Hallway Outside ST140/148)	

Poster Session 2 (Hallway Outside ST140/148)

2:00 – 2:30 pm	Odd Numbered Posters Present
2:30 – 3:00 pm	Even Numbered Posters Present
3:00 – 3:30 am	Refreshment Break / Group Photo (Hallway Outside ST140/148)

PI / Student Meetings (ST140/148)

3:30 – 4:30 pm	ST140 – PI Meeting
3:30 – 4:30 pm	ST148 – Student Meeting

Closing Keynote Lecture: Dr. Steven Jones (ST140)

Co-Director & Head, Bioinformatics, Genome Sciences Centre, British Columbia Cancer Research Centre

Chair: Dr. Athanasios Zovoilis

4:30 – 5:30 pm	Adventures in Long-read sequencing	**2**
5:30 – 7:00 pm	Seated Dinner and Awards Ceremony (TI160)	
7:30	Unofficial post-conference gathering (Kilkenny Irish Pub)	

2nd Annual BioNet Conference – Poster Presentations

1	**Lilit Antonyan, University of Calgary** Genome-Wide Association Study of Pediatric Obsessive-Compulsive Behaviors, the Correlation with Imaging Endophenotypes	**14**
2	**Reid McNeil, University of Calgary** Evaluating DNA Amplifications and their effect on Squamous Cell Carcinoma Prognosis using The Cancer Genome Atlas	**16**
3	**Adaobi Ojiakor, University of Alberta** The evolutionary diversification of the Salmonella artAB toxin locus	**17**
4	**Ana Nikolic, University of Calgary** Subclonal Copy Number Variation and Chromatin State Heterogeneity in Cancer Single-cell Chromatin Accessibility Data	**18**
5	**Rebecca Chen, University of Calgary** Modular amplicon deep sequencing pipeline for studying drug resistance resistance in parasites	**19**
6	**Vladimir Avramovic, University of Calgary** Use of reference population data for resolving variants of unknown significance in hematopoietic genes	**20**
7	**Afiya Razia Chida, University of Calgary** Deciphering structural variation in diverse natural genome isolates of *Caenorhabditis elegans*	**21**
8	**Deshan Perera, University of Calgary** CATE: A fast and scalable CUDA implementation to conduct highly parallelized evolutionary tests on large scale genomic data.	**22**
9	**Yashika Handa, University of Calgary** Quantifying the contribution of peer review to scientific publishing	**23**
10	**Qing Li, University of Calgary** Transfer-learning enables accurate variant prioritization and improves power in transcriptome-wide association studies: an application to breast cancer	**24**
11	**Catrione Lee, Agriculture and Agri-Food Canada, University of Lethbridge** Effect of Antimicrobial Use in Natural Versus Conventional Cattle Feedlots on the Microbiome and Resistome	**26**
12	**Sani-e-Zehra Zaidi, Agriculture and Agri-Food Canada, University of Lethbridge** Genomic characterization of Enterococcus hirae from beef cattle feedlots and associated environmental continuum	**27**
13	**Dinghao Wang, University of Calgary** cLD: Rare-variant disequilibrium between genomic regions identifies novel genomic interactions	**28**

14 **Luke Saville, University of Lethbridge** **29**
Processing of Alu and B2 SINE RNAs by Hsf1 is dysregulated in neuro-toxicity models and Alzheimer's disease patients

15 **Jiayi Bian, University of Calgary** **30**
An expression-directed linear mixed model (edLMM) discovering low-effect genetic variants

16 **Abhinaya Venkatesan, University of Calgary** **32**
Identification of a genetically divergent population of the human soil transmitted helminth Trichuris in Cote d'Ivoire that is unresponsive to albendazole-ivermectin combination treatment

17 **Eléonore Charrier, University of Calgary** **34**
Developing a modular resource to support comprehensive and flexible long-read and short-read rDNA metabarcoding across the nematode phylum

18 **Chenhua Li, University of Calgary** **36**
A transcriptomic view of the molecular mechanisms that the lancet liver fluke uses to manipulate its hosts behavior

19 **Elisabeth Richardson, University of Alberta** **37**
Microbiome analysis of tailings reclamation protocols in the Athabasca Oil Sands Region

20 **Marta Gerasymchuk, University of Lethbridge** **38**
Rejuvenation Effects of Phytocannabinoids and Anti-aging Drugs on Dermal Fibroblasts During Aging

21 **Ankita Narang, University of Calgary** **40**
Whole-genome sequencing analysis of clozapine-induced myocarditis

22 **Michael J Johnston, University of Calgary** **42**
Type B Ultra Long-Range Interactions in PFAs (TULIPs) represent recurrent epigenomic alterations in pediatric ependymoma

23 **Emmanuel Alimo, University of Alberta** **45**
Site of Antibody Affinity Maturation in Early Vertebrates

24 **(Christine) Xiaoli Liu, Agriculture and Agri-Food Canada Lacombe** **46**
PacBio Sequencing to Explore Microbiome Communities of Aquaculture Environments

25 **Venkateswara Rao Parimisetti, University of Calgary** **47**
Comparison of species taxonomy metabarcoding pipelines for bovine Eimeria COI and correlation with morphological species identification

26 **Liam Mitchell, University of Lethbridge** **48**
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

27 **Min Jae Kim, University of Calgary** **49**
Gene by environment interaction study of major depressive disorder and peer victimization in a pediatric population

28 **Riya Roy, University of Lethbridge** **50**
Identifying the impact of editing on B2 RNA stability: Connecting the dots between editing and cellular response to stress

29 **Daniel A. Salazar-Alemán, University of Calgary** **51**
A transcriptomics approach to understand the interactions between Escherichia coli and copper, silver, and gallium metal ions

30 **Travis Haight, University of Lethbridge** **52**
Dissecting the brain tissue specific expression of circular RNAs through Nanopore Sequencing

31 **Kaylee D. Rich, University of Calgary** **53**
Parasites and polyjuice: the role of molecular mimicry in modulating the host immune system

32 **Grace M. Mariene, University of Calgary** **54**
Performance assessment of long-read assemblers using nematode species with different genome natures

33 **Kevin Muirhead, University of Calgary** **55**
Species-specific prediction of glycogen-targeting enzyme activities in dominant vaginal *Lactobacillus* species

34 **Sandesh Acharya, University of Calgary** **57**
Stabilized Marker Gene Selection for single-cell RNA-seq data

35 **Kyle Lesack, University of Calgary** **59**
Input File Read Order Affects the Reproducibility of Structural Variation Calling from Long-Reads

36 **Rumika Mascarenhas, University of Calgary** **61**
Identifying somatic mutations from DNA derived from stereo-EEG electrodes in patients with focal cortical dysplasia

37 **Gregory Robinson, University of Lethbridge** **63**
Micropropagation and Transformation of *Cannabis sativa* L.

38 **Stephen M. J. Pollo, University of Calgary** **64**
Transcriptomic insights into the parasitic nematode *Heligmosomoides polygyrus*

39 **Megan Malach, University of Lethbridge** **65**
Investigating the Efficacy of Cannabis as a Novel Therapeutic for Aggressive Pediatric Brain and Nervous System Tumors

40 **Jason R. Grant, University of Alberta** **67**
Proksee: a web server for assembling, annotating, and visualizing bacterial genome

The 2nd Annual BioNet Conference
Keynote and Plenary Speakers

Dr. Fiona Brinkman, PhD, FRSC, Distinguished Professor, Simon Frasier University

Pleasing People and Pathogens: Bioinformatics Battles

Fiona Brinkman is a world-leading bioinformatics expert who has led research efforts, including large consortiums, to tackle the global health threats posed by infectious and inflammatory diseases and antibiotic resistance. Her research and widely used open-source computational tools have led to fundamental insights into how microbes evolve and are enabling health agencies to implement more sustainable control of infectious diseases and to preserve microbiota essential for human and environmental health.

Genomics and bioinformatics provide powerful methods that can aid such efforts and improve our understanding of microbial virulence, microbial evolution, and ourselves. The Brinkman laboratory comprises an interdisciplinary bioinformatics and "wet-lab" environment, investigating microbes and their diseases in the following ways:

1. Investigating the role in disease of both the microbe and its host (i.e immune system failure, including allergy and asthma), using genomics and systems biology-based approaches
2. Using genomics and network analysis to characterize disease outbreaks, their epidemiology, their evolution, and their environmental/social/genetic causes, and
3. Identifying new anti-infective and immune modulating therapies/biomarkers.

Our health, and the health of the environment and other animals, are interconnected. So we are also applying our approaches to "One Health" settings – applying our methods to aid environmental and animal health research efforts.

The overall goal is to capitalize on a combined bioinformatics and wet lab approach, integrating large datasets, to better understand key themes in infectious disease etiology and aid development of more sustainable approaches for disease management in a One Health context.

Affiliations:

Distinguished Professor, Department of Molecular Biology and Biochemistry, Simon Frasier University
Associate Professor, School of Computing Science and Faculty of Health Sciences, Simon Frasier University
Co-Lead, National IRIDA project and consortium
Co-Director, CIHR/MSFHR Bioinformatics Training Program
Co-Lead, Bioinformatics, IMPACTT National Microbiome Core
Computational Biology and Modelling Pillar Deputy, CoVaRR-Net
Core Faculty Member, Canadian Bioinformatics Workshops

Dr. Steven Jones, PhD, FRS(C), FCAHS, Director, Head, Bioinformatics, Distinguished Scientist, BC Cancer Research Institute

Adventures in Long-read sequencing

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

In 2005, Dr. Jones was identified as one of Canada's top 40 professionals under 40 by Caldwell Partners International as well as by Business in Vancouver. He has received the Spencer Award for IT innovation as well as the 2007 Medical Genetics teaching award from UBC. He is a founding director of the CIHR/MSFHR Bioinformatics Training Program as well as director of the UBC Bioinformatics Graduate Program. In 2011, he was inducted as a Fellow of the Royal Society of Canada for his contributions to Genomics and Bioinformatics, and in 2012 he was a recipient of the prestigious UBC Killam teaching prize in recognition of his contributions to graduate bioinformatic education. In May of 2014 Dr. Jones was awarded the Distinguished Achievement Award by the Faculty of Medicine at UBC and in June 2014 he became a Fellow of the Canadian Academy of Health Sciences. He was recognized by Clarivate Analytics in 2016 and 2018 as among the world's most highly cited researchers in his field.

Affiliations:

Canada Research Chair in Computational Genomics, University of British Columbia
Professor, Medical Genetics, University of British Columbia
Director Bioinformatics, Genome BC Bioinformatics Platform, Genome BC
Founding Director, CIHR/MSFHR Bioinformatics Training Program
Director, Bioinformatics Graduate Program, University of British Columbia
Professor, Genetics Graduate Program, University of British Columbia
Associate Member, Peter Wall Institute for Advanced Studies
Associate Member, Michael Smith Laboratories, University of British Columbia

Dr. Tim McAllister, Agriculture and Agri-Food Canada

Genomic approaches to unraveling the secrets of respiratory disease and antimicrobial resistance in beef cattle

Tim McAllister was raised on a mixed cow-calf operation in Inniisfail Alberta. He obtained his M.Sc. in Animal Biochemistry at the University of Alberta and his Ph. D. in microbiology and nutrition from the University of Guelph in1991. He is presently a principal research scientist with Agriculture and Agri-Food in Lethbridge, Alberta, Canada. Tim leads a diverse research team that has been studying antimicrobial resistance in beef cattle production systems since 1997. The team's recent work has focused on studying AMR from a "One Health" perspective using enterococci as AMR indicators in beef production, human sewage and clinical settings. He is also investigating the role of integrative conjugative elements in the transfer of antimicrobial resistance genes within the bacterial bovine respiratory disease complex. Tim has authored over 850 scientific papers, the recipient of several national and international awards and holds adjunct professorship appointments at several universities in Canada and abroad.

The 2nd Annual BioNet Conference
Invited Speakers

Dr. Pinaki Bose, University of Calgary

TBD

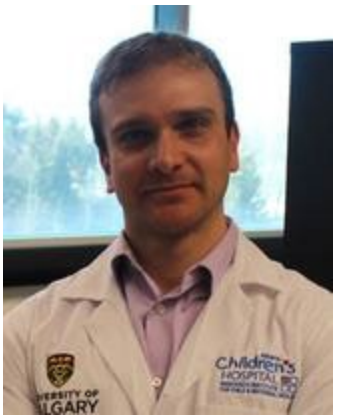
Dr. Pinaki Bose completed his Ph.D. in basic cancer research with Dr. Karl Riabowol at the University of Calgary, investigating the role of the ING1 tumour suppressor protein in DNA damage signaling and apoptosis. After completing his Ph.D., Pinaki joined the Ohlson Research Initiative (ORI) as a postdoctoral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

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

Dr. Jason de Koning, University of Calgary

The Population Genetics of Non-Equilibrium Molecular Evolution: A Unified Approach to Understanding Neutral and Non-Neutral Evolutionary Processes

Substantial analytical and computational challenges impede effective exploitation of the massive genomic data sets being generated by next-generation DNA sequencing technologies. To help overcome these impediments, my laboratory is developing novel bioinformatics approaches for the scalable, high-throughput, simultaneous analysis of many genomes using model-based statistical inference. We are particularly interested in bringing modelling advances in molecular evolution and comparative genomics to bear on interpreting personal genomic variation in humans.



Dr. Marco Gallo, University of Calgary

TBD

Cancer stem cells (CSCs) are rare cells required for tumor initiation and propagation in brain tumors. These cells are resistant to current therapies and constitute a reservoir that enables tumor recurrence. Our previous work has shown that CSCs are characterized by specific epigenomic states, which differentiate them from the other cells in the tumor bulk. Our goal is to define these CSC-specific epigenetic states and target them with next-generation epi-drugs.

Dr. Paul Gordon, University of Calgary

Is it signal or noise? Dissecting raw direct RNA data from nanopore devices

Paul Gordon, PhD is the Bioinformatics Manager for the Centre for Health Genomics and Informatics in the Cumming School of Medicine at the University of Calgary. Dr. Gordon began working in bioinformatics in 1996, part of the team sequencing the first complete archaeal genome at the National Research Council of Canada. Subsequently he has employed computational methods for the analysis of everything from fungal genomes for biofuel enzyme discovery, to organ transplant rejection detection, cancer immunology, genomic epidemiology, and rational oligonucleotide design, with 58 peer-reviewed papers. He has been working with nanopore data since the first North American case of high pathogenic avian influenza in early 2014, developing and maintaining a number of open-source tools for nanopore data acquisition and analysis since the time.

Dr. Devin Holman, Agriculture and Agri-Food Canada

The swine gut microbiome and resistome in conventional and organic production systems

Dr. Devin Holman is currently a Research Scientist in Livestock Microbiology at Agriculture and Agri-Food Canada based at the Lacombe Research and Development Centre (RDC) in Lacombe, AB. He has a M.Sc. degree in Food Microbiology from Dalhousie University in Halifax, NS, and a Ph.D. in animal science and microbiology from McGill University in Montreal, QC. Dr. Holman was a postdoctoral fellow with AAFC at the Lethbridge RDC focusing on the nasal microbiome and its association with bovine respiratory disease in beef cattle. He also did a postdoctoral fellowship with the USDA-Agricultural Research Service at the

National Animal Disease Center in Ames, IA, where his research involved challenging pigs with multidrug-resistant *Salmonella Typhimurium* and investigating the effect that high-dose antimicrobial administration had on *S. Typhimurium* shedding and on the pig gut microbiome. Since starting with AAFC in Lacombe in June 2017, Dr. Holman has been involved in several swine projects designed to characterize the transfer of antimicrobial resistance genes among pigs and to determine the effect of various management and antimicrobial treatment strategies on the pig gut microbiome. Overall, his work is largely focused on monitoring antimicrobial resistance in food producing animals, reducing antimicrobial use, and improving animal health and performance through microbiome-driven research.

Dr. Gane Ka-Shu Wong, University of Alberta

Drug targets and lead compounds in the era of "free" data for humans and biodiverse taxa

Professor Gane Ka-Shu Wong was a founding member of two internationally renowned genomics research organizations. The first was at the University of Washington. The other, named Beijing Genomics Institute (BGI or 華大基因), subsequently grew into one of the largest such ventures in the world and hosted a successful IPO in 2017. His h-index is 74, with a total of 37743 citations, and 3840 citations/year in 2021 (Google Scholar, 06 May 2022). He has 38 publications in top-tier journals like Nature Publishing Group and Science. In the 2019/2020/2021 rankings at Clarivate Analytics, he is a Highly Cited Researcher for the Cross Field category with 20 papers since 2012 rated in the top 1% by citations for their respective fields. He is currently a University of Alberta professor, jointly appointed in the Faculty of Medicine and the Faculty of Science.

His research lies at the cutting edge of technologies, high throughput sequencing and computer analyses, that are revolutionizing biology and medicine. Many of these efforts are multidisciplinary international consortiums, which he leads. These include, from the early days of BGI, landmark papers on genomes for indica rice, silkworm, and chicken. More recently, he led the 1KP initiative that sequenced the transcriptomes for a thousand phylodiverse species of green plants (Oct 31 2019 cover of Nature, 2020 Annual Reviews of Plant Biology). He now leads a successor project called 10KP, sequencing complete genomes for 10x as many species. 1KP was a "first-of-kind" project (for all taxa) when launched. It inspired similar projects for other taxa, leading ultimately to the Earth BioGenome Project now sequencing all known eukaryotic species. It also discovered several novel optogenetic proteins (2014 Nature Methods) that are widely used for studies of the mammalian brains in awake behaving animals. Many licenses have been signed, most notably with GenSight Biologics, whose method to restore vision in patients with retinitis pigmentosa was granted fast-track designation by US-FDA regulators in late 2021.

Another component of his research is the introduction of omics technologies into problems of interest for medicine. These include bacterial metagenomics to monitor patient response in fecal microbial transplantation (2017 JAMA, 2016 Frontiers in Microbiology), proviral integration into

genomes of patients with autoimmune disease of suspected infectious etiology, single-cell omics to elucidate metastasis progression in breast cancer (2021 Cell Reports Medicine), and novel nanotechnologies for sequencing low abundant DNA/RNA in clinical samples where the target is below PCR limits of detection (2021 Scientific Reports). With the rapid growth in digital health databases, and the breakthroughs in deep learning algorithms, he has been charting a new path in drug target discovery that will increase the likelihoods of success for clinical trials (2021 npj Genomic Medicine) by eliminating our reliance on animal models and cell lines.

Dr. Wong's diverse interests are a reflection of his life history. Originally born in Hong Kong, he emigrated to Canada at the age of six, and grew up having to navigate two cultures. Scientifically, he did not start in biology or medicine. His B.A.Sc. at the University of British Columbia included honours in mathematics, physics, electrical engineering, and computer science. His Ph.D. studies at Cornell University were on experimental low temperature physics. While a postdoctoral fellow at the California Institute of Technology, he was recruited to the nascent Human Genome Project and moved to the University of Washington. There, he met the key people who would eventually form BGI, and some of the people he mentored at BGI have since launched their own companies. He admits that luck played a major role in his career successes.

Dr. Igor Kovalchuk, University of Lethbridge

Transgenerational response to heat in Arabidopsis thaliana - genetic and epigenetic components

Igor Kovalchuk is a plant biologist with the focus on genetics and epigenetics of stress response and transgenerational stress tolerance. He is interested in the role of epigenetics in adaptation and microevolution.

Dr. Olga Kovalchuk, University of Lethbridge

TBD

Olga Kovalchuk studies the role of epigenetic dysregulation in carcinogenesis and regulation of the cancer treatment response. Her work also focuses on radiation epigenetics and the role of epigenetic changes in genome stability, carcinogenesis oncogenic signaling, DNA damage, repair, and recombination.

Dr. Ian Lewis, University of Calgary

Predicting the severity of infections via microbial proteomics: insights from a city-wide survey of blood stream infections over 20 years

Infectious diseases account for the majority of deaths in developing countries and are among the ten leading causes of mortality in Alberta. The emergence of diverse drug resistant pathogens has made finding new antimicrobial therapies both a Canadian and global health priority. My laboratory investigates the connection between metabolic adaptation and virulence of human pathogens. The ultimate goal of this research is to develop new diagnostic methods to identify high risk patients and novel antimicrobial therapies to control infections. One of the key challenges my laboratory faces is unraveling the complex host-pathogen metabolic dynamics that occur during infections. To address this challenge, we use state-of-the-art mass spectrometry (MS) and nuclear magnetic resonance (NMR) technology to comprehensively quantify the flow of molecules between pathogens and their hosts. This metabolomics approach is a powerful strategy for connecting metabolic phenomena with specific genes and is essential for understanding the metabolic determinants of virulence.

Dr. Quan Long, University of Calgary

Discovering genetic basis of complex traits by utilizing "borrowed data-bridge

Quan Long, PhD, was trained in both mathematics and computer science. His current research focuses on phenotype predictions and gene mapping via integration of multiscale omics using statistics and informatics. He is also interested in within-host evolutionary analysis using sequences data that may lead better prediction of disease status such as cancer progression and pathogen antimicrobial resistance. He was a staff R & D engineer analyzing memory leak at IBM Research; then a staff scientist serving for the 1,000 Genomes Project and other evolution-focused projects at the Wellcome Trust Sanger Institute. Afterwards, he assumed the position of a postdoc fellow at the Gregor Mendel Institute, working on methods development as well as real data analysis for NGS-based variants calling, association mapping, and population genetics. Before joining University of Calgary, he was an assistant professor (research track) in Icahn School of Medicine at Mount Sinai, working on phenotype predictions and gene expression networks

Dr. Tarah Lynch, Alberta Precision Laboratories

COVID-19: A Catalyst for Change

Tarah Lynch earned a Ph.D in Microbiology and Infectious Diseases from the University of Calgary studying host-pathogen interactions of Vibrio parahaemolyticus. She continued her training at the Centers for Disease Control and Prevention in Fort Collins, Colorado through an ASM/CCID Fellowship where her focus was on Bartonella detection and assay development. Dr. Lynch then transitioned into bioinformatics through her work in bacterial genomics of Clostridium difficile at the Public Health Agency of Canada's National Microbiology Laboratory.

Dr. Lynch currently holds a position at the Alberta Provincial Public Health Laboratory (south site) as a Program Lead for Genomics and Bioinformatics where her work focuses on infectious disease surveillance and diagnostic test development using high throughput sequencing technologies and bioinformatics.

Dr. Gerlinde Metz, University of Lethbridge

Can't buy me love: epigenetic and metabolomic determinants of health linked to social isolation

Dr. Gerlinde Metz is a Professor of Neuroscience and a Tier 1 Board of Governors Research Chair (Healthy Futures) at the Canadian Centre for Behavioural Neuroscience at the University of Lethbridge. She also is an Adjunct Professor with the Department of Obstetrics and Gynecology, Faculty of Medicine & Dentistry, at the University of Alberta. Her research program focuses on the influence of environmental factors on behaviour, brain health and disease. Her laboratory has developed unique models to find that stress affects lifetime health of future generations. Her ground-breaking studies of transgenerational programming of disease are now leading to the discovery of new biomarkers for risk prediction and early diagnosis of disease. Moreover, by working with families affected by natural disaster, war and other tragedies, Dr. Metz's team aims to develop effective strategies that will reduce the impact of stress and build resilience.

Rodrigo Ortega-Polo, Agriculture and Agri-Food Canada

The Computational Analyses Research Support Unit of the Lethbridge Research and Development Centre

Rodrigo Ortega Polo has a Bachelor of Science and a Master of Science in Biological Sciences from the University of Lethbridge in Alberta, Canada. He worked at the Lethbridge Laboratory of the National Centres for Animal Disease of the Canadian Food Inspection Agency (CFIA) from 2012-2017. In his role with the CFIA he performed the bioinformatic design of the development of molecular tests for detection and subtyping of high-consequence animal viruses such as foot-and-mouth disease virus and avian influenza virus. The virology research group in Lethbridge then became the genomics unit of the National Centre for Foreign Animal Disease in Winnipeg where Rodrigo worked on setting up computers, performed comparative genomics analyses, and streamlined workflows for data analyses. After that experience, Rodrigo was appointed as project manager/bioinformatics analyst at the University of Lethbridge with Dr. Erasmus Okine (who was then the Vice-President Research) on a project funded by the Agricultural Greenhouse Gases Program of Agriculture and Agri-Food Canada (AAFC). Rodrigo was in charge of a budget of $1.2 million for the 5-year project, and coordinated the recruitment and appointment of technical staff, postdoctoral fellows, and summer students. He also coordinated the efforts with collaborators from other universities and the federal and provincial governments. In that same position he used his bioinformatics skills to collaborate with federal government scientists on studies about the microbial communities and the collection of antimicrobial resistance genes across the One Health continuum.

Rodrigo has been the Biology Study Leader - Bioinformatics at the Lethbridge Research and Development Centre of AAFC since September 2018. In this capacity, he is the team lead of the Computational Analyses Research Support Unit. The unit has grown considerably since the unit started with only one person in 2018. Today, the Lethbridge CARSU is the largest AAFC computational support

unit in the regions. The unit leads a Genome Alberta – Alberta Innovates EBS project, and collaborates on several projects including a Genome Canada – AAFC Large Scale Applied Research Project on honey bee health called BeeCSI.

The Lethbridge CARSU is diverse in background and expertise, including omics (i.e. genomics, metagenomics, transcriptomics), image analysis and phenomics, application development and cloud computing. Besides training AAFC staff and graduate students who do their research with the federal government, several co-op and term students have had the opportunity to work at CARSU and train to become the bioinformaticians and data analysts of the future. Lethbridge CARSU members have been co-authors in thirty peer-reviewed research publications since 2018.

Rodrigo is also a member of the AAFC Bioinformatics Research Support Network (BRSN), a national team of experts from the Science and Technology Branch and the Information Systems Branch who work together to provide bioinformatics support and training to science programs across the organization.

Dr. Paul Stothard, University of Alberta

Bioinformatics for genetic improvement, management, and conservation of livestock

Dr. Stothard is a Professor at the University of Alberta in animal molecular biology and bioinformatics. His research is focused on the genetic basis on important traits in animals, with the use of whole genome sequencing and bioinformatics to understand the genetic bases of production and health traits, including pathogen resistance and resilience, in livestock.

Dr. Laura Sycuro, University of Calgary

Piercing the fog of the gut-brain axis with metagenomics

Dr. Laura Sycuro is an Assistant Professor at the University of Calgary and Genomics and Bioinformatics Theme Lead for the International Microbiome Centre. The broad goal of her research program is to harness the microbiome to promote the health of women and children. To this end, Dr. Sycuro's lab couples microbiome discoveries relating to infection, preterm birth and neurodevelopment with the work of her social enterprise targeting gender-based health inequities.

Dr. Maja Tarailo-Graovac, University of Calgary

Genome sequencing and complex disease mechanisms in rare disease patients

Dr. Maja Tarailo-Graovac is an Associate Professor at Cumming School of Medicine in the departments of Biochemistry & Molecular Biology and Medical Genetics. Her research program interconnects high throughput sequencing, human and model organism genetics/genomics to better detect/interpret pathogenic variants and understand effects of different genetic backgrounds (i.e. genetic modifiers) on variable phenotypic outcomes in rare genetic disease.

Dr. Edwin Wang, University of Calgary

Predictive genomics, from understanding of genomics to predictive models

We are conducting both computational and experimental systems biology toward precision medicine. The computational work includes: (1) big medical data analysis (2) machine learning and deep learning (3) predictive model construction based on genomic data of diseases including cancer. We are developing novel algorithms for modeling of molecular networks and cancer biomarker discovery, and also developing new concepts for data analysis toward interpreting data, generating, prioritizing and testing new hypotheses.

Dr. James Wasmuth, University of Calgary

A hitchhiker's guide to a parasite's genome

Dr. James Wasmuth is a biochemist who hated the wet lab and fell in love with Perl and BLAST. He earned his degrees at Imperial College, and the Universities of York and Edinburgh, and completed his postdoctoral training at the Hospital for Sick Children in Toronto. Now at the University of Calgary's Faculty of Veterinary Medicine, Dr. Wasmuth and his team use bioinformatics, genomics, transcriptomics, and metabolomics to understand how parasites successfully infect their hosts. His lab is funded by NSERC, RDAR, and Alberta Environment and Parks to identify new drugs and vaccines against parasites of livestock and fish. To the chagrin of his students and postdocs, he still codes in Perl. In 2019, Dr. Wasmuth has appointed as Director of the Graduate College, a pan-Faculty community that strives to add value and leadership experience for UCalgary graduate students and postdocs.

Dr. Athanasios Zovoilis, University of Lethbridge

NERD-seq a novel approach of Nanopore direct RNA sequencing

Dr. Zovoilis is an Associate Professor of Bioinformatics at the University of Lethbridge, which he joined from Harvard Medical School (USA) as a Canada Research Chair in RNA Bioinformatics and Genomics. He is the Director of the Southern Alberta Genome Sciences Center (SAGSC) as well as director of the SAGSC bioinformatics platform. He is also an adjunct member of the Canadian Centre for Behavioral Neuroscience. He is the academic lead of the Alberta Bioinformatics Network (BioNet), a Genome Canada supported initiative among University of Alberta, University of Calgary and University of Lethbridge. Dr Zovoilis is a physician with a background in Medical Genetics, a doctorate in Human Genetics from the University of Goettingen (Germany), postgraduate training in bioinformatics from the University of Manchester (UK), and expertise in bioinformatics of next generation sequencing from his time as research fellow at Vancouver Genome Sciences Centre (Canada) and Harvard Medical School (USA). He has also been a fellow of the European Molecular Biology Organization; a fellow of the German Research Foundation and he is also member of the Canadian Centre for Behavioral Neurosciences. Dr. Zovoilis' research combines translational research in medicine, in particular aging associated diseases such as dementia, with basic research in RNA genomics and bioinformatics. His competence in cutting edge technologies such as next-generation sequencing (NGS) is demonstrated by multiple leading or senior author scientific publications in distinguished journals such as Cell, Science, Elife, PNAS, EMBO and EMBO Rep. Being the Academic Lead of BioNet Alberta, he works to foster research collaborations and partnerships in the fields of bioinformatics and computational biology in Alberta and across Canada.

The 2nd Annual BioNet Conference Presentations

Genome-Wide Association Study of Pediatric Obsessive-Compulsive Behaviors, the Correlation with Imaging Endophenotypes

Lilit Antonyan[1], S-M Shaheen[1], David R Rosenberg[2], Gregory L Hanna[3], Paul D Arnold[1]

[1]University of Calgary, Calgary AB, Canada
[2]Wayne State University, Detroit, MI, USA
[3]University of Michigan, Ann Arbor, MI, USA
Presenting author: lilit.antonyan1@ucalgary.ca

KEYWORDS

Obsessive-Compulsive, GWAS, Quantitative, Endophenotypes, Imaging

ABSTRACT

Introduction: Obsessive-compulsive (OC) behaviors (OCB), characterized by intrusive thoughts and repetitive behaviors are common in youth [1]. Here, we examine the genetic basis of OCB, as well as their association with neuroimaging endophenotypes. The overall goal is to determine the relationship between brain activity and childhood OCB within the Research Domain Criteria (RDoC) paradigm [2]. Consistent with the RDoC, the Obsessive-Compulsive Scale of the Child-Behavior Checklist Scale (CBCL-OCS) is used as a quantitative measure [3].

There are a vast number of studies reporting the involvement of the cortico-striato-thalamo-cortical (CSTC) model in the pathophysiology of OCD [4]. In this study, cortical thickness (CT) and myelin water fraction (MWF) will be used as neuroimaging measures to delineate imaging endophenotypes for OCB.

Methods: Genotyping analysis was performed on 863 non-related pediatric subjects, including OCB cases and healthy controls. Four different genome-wide arrays (~3million single nucleotide variants (SNVs)) were used. Our multi-step genome-wide association study (GWAS) will identify statistically significant SNVs. Quality control (QC) and association analysis were conducted using PLINK software [5]. Later, SNVs were imputed via Minimac3 [6]. Consequently, post-GWAS analyses are in process. First, polygenic risk score analysis will be performed to test whether polygenic risks from larger GWAS of OC traits are associated with OCB in our dataset [7]. Second, SNV-based and gene-based analyses will be carried out. Further, imaging data analysis will be conducted to identify association of OC traits and define putative neuroimaging endophenotypes. Imaging data (CT and MWF) was collected from 3-dimensional T1 and T2 weighted structural scans.

Results: After QC, 627 samples with corresponding CBCL-OCS scores, and 1.8 million markers passed the QC filters. Although none of the SNVs passed the p-value threshold for genome-wide significance, a few markers were observed to be close to significance. Top variants were identified for further analysis. The top ranked loci associated with CBCL-OCS scores as a quantitative trait are identified. Initial SNV-based

and gene-based analyses showed that KCTD8(4p13) and DENND5B(12p11) genes are differentially expressed in specific brain tissues.

Moreover, we have already obtained CT and MWF measures for ~200 individuals who participated in our genetic study. Analyses of imaging endophenotypes are in progress.

Conclusion: To our knowledge, this project will be the largest to date to report genetic markers, and imaging endophenotypes of susceptibility to OCB in a pediatric-clinic-based population.

It will provide insight into the pathophysiology of OC traits mediated by structural and functional changes in specific brain regions.

REFERENCES

1. https://pubmed.ncbi.nlm.nih.gov/8083134/
2. https://apps.who.int/iris/handle/10665/41842
3. https://pubmed.ncbi.nlm.nih.gov/15335349/
4. https://pubmed.ncbi.nlm.nih.gov/24840803/
5. https://pubmed.ncbi.nlm.nih.gov/17701901/
6. https://pubmed.ncbi.nlm.nih.gov/22820512/
7. https://pubmed.ncbi.nlm.nih.gov/29132412/

Evaluating DNA Amplifications and their effect on Squamous Cell Carcinoma Prognosis using The Cancer Genome Atlas

Reid McNeil[1], Faisal M. Khan[1], Pinaki Bose[1]
[1]University of Calgary, Calgary, Canada

Presenting Author: reid.mcneil1@ucalgary.ca

KEYWORDS
TCGA, Bioinformatics, Squamous Cell Carcinoma, DNA Amplifications, Prognosis

ABSTRACT
The Cancer Genome Atlas (TCGA) has analyzed 11,000 tumors from 33 cancers, generating over 2.5 petabytes of genomic, transcriptomic, proteomic, and epigenomic data. Among these, comprehensive characterization of ~1800 squamous cell carcinomas (SCCs) including bladder (BLCA), esophageal (ESCA), cervical (CESC), lung (LUSC) and head and neck (HNSC) cancers has led to the discovery of squamous cell-specific genomic aberrations governing carcinogenesis and progression.

Few targeted therapeutics are used in the clinical management of SCCs and survival rates for these cancers remain dismal. Most recurrent mutations observed in SCCs are loss-of-function alterations in tumour suppressors that are difficult to target therapeutically. SCCs harbor commonly shared somatic copy number alterations (sCNAs), one form of which is the amplification of certain regions of DNA. DNA amplifications have been shown to trigger overexpression of multiple genes. Our objective was to identify driver genes connected to poor prognosis in regions of DNA amplification across different SCC sites in the TCGA database.

We performed integrative analyses of multi-omic TCGA datasets and devised a computational pipeline to identify drivers of oncogenesis. We used the GISTIC 2.0 pipeline to identify common genomic regions significantly amplified across the TCGA Pan-SCC cohort. Next, DESeq2 was used to identify significantly differentially expressed (DE) mRNAs and mircoRNAs within these amplified regions. DE analysis was also performed between cancer and matched normal samples. Survival analysis identified DE genes with prognostic value. Genes of prognostic value were collected into genesets and input for GSVA enrichment analysis to develop signatures of DNA amplifications in SCCs. Ingenuity Pathway Analysis was used to identify upstream regulators, and enrichment in biological processes and pathway.

GISTIC 2.0 confirmed several recurrent regions of DNA amplification across all SCC sites. DESeq2 identified ~50-200 significantly DE mRNAs and ~10-50 microRNAs per SCC site in amplified regions including 3q22-29, 5p15.33-15.1, 8q22.2-24.3, and 11q13.1-13.4. We identified genes with known oncogenic and tumor suppressor functions, and several others whose connection to cancer biology remains unknown. Survival analysis shortlisted mRNAs/microRNAs with high expression levels that correlated to worse patient survival. Gene Set Variation Analysis (GSVA) enrichment analysis scored samples for an “amplification signature” and higher scores correlated to poorer survival. Our work provides an overall picture of how DNA amplifications influence cancer biology and clinical presentation and revealed vulnerabilities that could help in developing new personalized treatments for SCCs.

The evolutionary diversification of the Salmonella artAB toxin locus

Adaobi Ojiakor[1], Rachel N. Gibbs[1], Zhe Chen[2], Xiang Gao[2] and Casey C. Fowler[1]

[1]University of Alberta, Edmonton, Canada
[2]Shandong University, Qingdao, China

Presenting author: ojiakor@ualberta.ca

KEYWORDS
Bacterial pathogenesis; *Salmonella* Typhi; typhoid fever; *Salmonella* Typhimurium; bacterial toxins; AB5-type toxins; typhoid toxin; ArtAB toxin; toxin evolution

ABSTRACT
Salmonella enterica is a diverse species of bacterial pathogens comprised of over 2500 serovars with variable host ranges and virulence properties. A number of literature indicate that two AB_5-type toxins, typhoid toxin and ArtAB toxin, play important roles in the pathogenesis and virulence outcomes of the *Salmonella* strains that encode them. Recently, two distinct types of *artAB*-like genetic elements were uncovered in *Salmonella* and they include those that encode ArtAB toxins (*artAB* elements) and those in which the *artA* gene is degraded and the ArtB homolog, also known as PltC, serves as an alternative delivery subunit for typhoid toxin (*pltC* elements). Using a multifaceted approach with bioinformatic tools at the core of it, this presentation will highlight the evolutionary diversification of the *artAB*-like genetic elements in various *Salmonella* serovars. We have identified 7 subtypes of ArtAB toxins that are encoded by 11 different groups of prophages within the *Salmonella* genus, emphasizing the key role that bacteriophages play in the evolutionary expansion of *artAB*-like genetic elements. Additionally, genetic and structural analysis has revealed important features that differentiate *pltC* elements from *artAB* elements. We have also identified evolutionary adaptations that allow typhoid toxin's PltC to efficiently interact with the other subunits that make up this remarkable toxin. For both *pltC* and *artAB*, we see that the B subunit sequences have undergone considerable evolutionary diversification, especially in regards to the amino acid residues that fine tune the chemical environment of their glycan binding pockets. This study provides a framework to understand the remarkably complex collection of *Salmonella artAB*/*pltC*-like genetic elements and provides in a window into the mechanisms of evolution for AB_5-type toxins.

Subclonal Copy Number Variation and Chromatin State Heterogeneity in Cancer Single-cell Chromatin Accessibility Data

Ana Nikolic[1,2,3], Divya Singhal[1,2,3], Katrina Ellestad[1,2,3], Michael Johnston[1,2,3], Yaoqing Shen[4,5], Aaron Gillmor[1,2,3], Sorana Morrissy[1,2,3], J. Gregory Cairncross[1,6], Steven Jones[4,5], Mathieu Lupien[7,8,9], Jennifer A Chan[1,2,6], Paola Neri[1,6], Nizar Bahlis[1,6], Marco Gallo[1,2,3]

[1]Arnie Charbonneau Cancer Institute, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada
[2]Alberta Children's Hospital Research Institute, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada
[3]Department of Biochemistry and Molecular Biology, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada
[4]Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, BC, Canada
[5]Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada
[6]Department of Oncology, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada
[7]Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada
[8]Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada
[9]Ontario Institute for Cancer Research, Toronto, ON, Canada

Presenting author: nikolica@ucalgary.ca

KEYWORDS
Epigenetics, single-cell epigenetics, bioinformatics, cancer biology, tumour heterogeneity

ABSTRACT

Clinical cancer specimens are highly heterogenous mixtures of neoplastic and stromal cells. Single-cell sequencing technologies enable high-resolution profiling of individual cells within these complex specimens. In ATAC-seq experiments, and accurate delineation of tumour and non-tumour cells can be challenging in many tumour types, due to a lack of established cell type markers for many cancers and uncommon microenvironmental cell types. Copy number alterations and focal amplifications are common alterations seen in many different cancer types, and represent one potential way to distinguish tumour from non-tumour cells in single-cell experiments.

We have developed an R package, Copy-scAT, which identifies areas of putative extrachromosomal amplification and large-scale CNVs in single-cell ATAC-seq experiments. We validate our method using datasets from adult glioblastoma, pediatric high-grade glioma, and multiple myeloma. Moreover, we show that the presence of copy number alterations significantly influences sample clustering and differential accessibility. In addition, we explore the influence of subclonal genetic and epigenetic phenotypes in adult glioblastoma, and identify intratumoral epigenetic differences at the subclonal level. Overall, we show that Copy-scAT[1] enables improved delineation of neoplastic and non-neoplastic cells and clonal dynamics in single-cell accessibility cancer datasets.

REFERENCES

1. https://doi.org/10.1126/sciadv.abg6045

Modular amplicon deep sequencing pipeline for studying drug resistance resistance in parasites

Rebecca Chen[1], Michaela Seal[1], Camila Queiroz[1], Lynsey Melville[2], Elizabeth Redman[1], Russell Avramenko[1], Dave Bartley[2], and John Gilleard[1]

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

Presenting author: pcrchen@ucalgary.ca

KEYWORDS
bioinformatic pipeline, molecular epidemiology, drug resistance, next-generation sequencing, parasitic nematode

ABSTRACT

Routine 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.

New DNA sequencing technologies are revolutionizing our ability to undertake surveillance and study how drug resistance emerges and spreads, but the application of these approaches to anthelmintic drug resistance in nematodes is still in its infancy. This is partly due to a lack of easily usable tools to analyze these large DNA sequencing data sets for molecular epidemiological studies.

We have therefore developed a modular and flexible bioinformatic pipeline written in R programming language to study the molecular epidemiology of anthelmintic drug resistance. This pipeline takes raw short-read amplicon sequencing data and applies several different modules to process the data. The results are then visualized via haplotype networks and geospatial maps. This pipeline has been designed with flexibility in mind to allow different molecular markers to be easily incorporated as our molecular understanding of anthelmintic resistance emerges and spreads. The modular design of the pipeline allows a review-as-you-go workflow to allow higher quality control.

We have used benzimidazole in small ruminant parasitic nematode species to develop the pipeline as several important resistance mutations in the isotype-1 β-tubulin drug target are well characterized. We illustrate its use on two different large-scale isotype-1 β-tubulin amplicon sequencing data sets from sheep nematode populations to compare early-stage (*Nematodirus battus* from 150 UK sheep farms) and late-stage (*Haemonchus contortus* from 130 North American sheep farms) benzimidazole resistance. The haplotype networks and geospatial maps illustrate the localized emergence of resistance in its early stages and the more complex patterns in late-stage resistance suggestive of derivation from multiple origin regions in the late-stage population. This toolkit should provide a valuable resource for molecular epidemiology analysis of anthelmintic drug resistance educational and communication activities and is adaptable to molecular epidemiological studies of other organism groups.

Use of reference population data for resolving variants of unknown significance in hematopoietic genes

Vladimir Avramovic[1], Catherine Diao[1], Tatiana Maroilley[1], Maja Tarailo-Graovac[1]
[1] University of Calgary, Calgary, Canada

Presenting author: vladimir.avramovic@ucalgary.ca

KEYWORDS
Rare disease, genome sequencing, variant interpretation, hematopoietic genes, reference population

ABSTRACT
Despite the significant improvement that genome sequencing (GS) brought to rare disease diagnostics, the average diagnostic rate is only around 40%[1]. One of the major obstacles for further improvement of GS diagnostic pipelines efficiency is still limited ability for interpretation of the sequencing data[2]. While this applies in general, the interpretation of certain variant types is especially difficult. This group includes synonymous[3], intronic[4] and UTR[5] variants, among others.

Here we present a strategy to improve the interpretation of variants in hematopoietic genes that are currently classified as "variants of unknown significance" (VUS). Currently, germline variants in approximately 50% from around 160 known hematopoietic genes have been associated with rare diseases[6,7]. Many of them have been also associated with cancer, as somatic[7].

Previously, we showed that somatic variants from dozens of hematopoietic genes implicated in human rare diseases may be present in the reference population datasets[7]. These variants often drive the process of clonal expansion, by which they become more abundant in the blood tissue and subsequently appear in reference population datasets misclassified as germline variants. We proposed that the presence of these somatic variants in population databases[8,9] can be used as a feature in combination with their potential relation to cancer[10] and the sequence conservation scores[11] for prediction of their functionality in rare diseases, in germline. For the preliminary assessment, we considered the variants from the genes already known to be associated with rare diseases, *ASXL1* and *DNMT3A.* The results we obtained with *ASXL1* nonsense and *DNMT3A* missense VUS (the types known to cause severe pediatric rare diseases, Bohring Opitz syndrome and Tatton-Brown syndrome, respectively) confirmed our basic hypothesis. Almost all of the *ASXL1* nonsense variants and around 30% of *DNMT3A* missense variants from TOPMed[8] database show signs of somatic origin, while their assessment with a selected combination of *in silico* pathogenicity predictors[12,13] returned high pathogenicity scores. By expanding the assessments to include synonymous, intronic and 5'UTR VUS, which can potentially cause similar consequences by splice-altering mechanisms (e.g. protein truncation), we obtained a list of candidate variants that we are now experimentally validating. We expect that at least some of these variants with evidence of somatic mosaicisms in population databases will be shown to alter gene function. The experimental confirmations would further support evidence of somatic mosaicism in untargeted population as evidence of detrimental effect and driver potential of the variants in hematopoietic genes currently classified as VUS.

Deciphering structural variation in diverse natural genome isolates of *Caenorhabditis elegans*

Afiya Razia Chida[1,2], Victoria Rodrigues Alves Barbosa[1,2], Tatiana Maroilley[1,2], Andrew Galbraith[1,2], Filip Cotra[1,2], Maja Tarailo-Graovac[1,2]

1. Departments of Biochemistry, Molecular Biology and Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada T2N 4N1, Canada
2. Alberta Children's Hospital Research Institute, University of Calgary, Calgary, AB T2N 4N1, Canada

Presenting author: afiya.chida@ucalgary.ca

KEYWORDS

Structural Variants, Natural isolates, Short Reads Sequencing, Genetic Diversity, Hypervariable Regions

ABSTRACT

Genetic variation can be comprehended in an effective manner with thorough knowledge on both Single Nucleotide and Structural Variants (SVs). Whilst the impact of SVs is quite prominent, research involving them is sparse. Genomic rearrangements such as Copy Number Variations, Inversions as well as translocations contribute to about 34% of known disease-causing variants in humans [1]. Since the characterization and annotation of SVs in humans is arduous, we use *Caenorhabditis elegans* model organism. Despite the current view that Short Read Whole Genome Sequencing (SR-WGS) can be a major limitation in identifying SVs, we have experimentally validated and confirmed their presence in *C. elegans* balancer strains – strains with chromosomal rearrangements that maintain heterozygosity and prevent crossover events by balancing large regions of genome [2]. To extend the diversity beyond mutagenized balancer strains, generated in the N2 (Bristol, UK) strain background, we also use *C. elegans* strains from five diverse natural backgrounds such as CB4856 (Hawaii, USA), JU1400 (Seville, Spain), AB1 (Adelaide, Australia), GXW1 (Wuhan, China) and KR314 (Vancouver, Canada). We performed SR-WGS of the strains followed by advanced bioinformatics analysis using combination of tools with specialized in-house methodologies. We identified SVs of variable sizes as well as complexity. We analyzed these based on different criteria like size, location, impact on the coding region and their uniqueness among the strains. Selected candidate SVs were experimentally validated using PCR. Unlike the balancer strains advantaged by the excellent N2-based reference genome, we faced challenges with some of the breakpoints in the natural isolate strains that occurred in the regions of high variability. Such breakpoints were identified mainly on the left-arm of LGII and right-arm of LGV. These highly divergent regions have an unfavorable impact on the characterization and validation of SVs. Though we were successful in identifying most of the SVs using SR-WGS, we hypothesize that these regions of high variability seen across the strains could be due to the lack of population specific reference [3] and thus the SVs within those regions may be resolved by utilizing either long read sequencing or an adequate reference genome. Finally, our attempt to capture the spectrum of variants in natural isolate strains aids in better understanding of genetic and phenotypic heterogeneity due to variable genetic backgrounds. The knowledge gained from our *C. elegans* genomics research is being successfully translated to human genomics by developing tractable strategies in improving the calling and interpretation of SVs in humans.

CATE: A fast and scalable CUDA implementation to conduct highly parallelized evolutionary tests on large scale genomic data

Deshan Perera[1], Christian Hubber[2], Quan Long[1]

1. Department of Biochemistry & Molecular Biology, Cumming School of Medicine, University of Calgary, Calgary, AB T2N 4N1, Canada.

2. Department of Biology, The Pennsylvania State University, University Park, 16802 PA, USA

Presenting author: duwagedahampriyabala@ucalgary.ca

KEYWORDS
CUDA, Molecular evolution, Neutrality tests, 1000 Genomes Project, Multiprocessing

ABSTRACT
[BACKGROUND]
Algorithm parallelization has proven to reduce computational time. The Graphical Processing Unit (GPU) housing thousands of cores enables the potential of large-scale parallelization. NVIDIA's CUDA based GPUs are becoming commonplace in both general purpose and specialized computing. Presently, the use of parallel processing to solve genetic algorithms with the aim of reducing computation time has gained traction. However, so far, such potential of high parallelization has not been realized in molecular evolution analyses.
[METHODS]
We have built a scalable program using NVIDIA's Compute Unified Device Architecture (CUDA) platform together with an exclusive file hierarchy to process six different tests frequently used in molecular evolution, namely: Tajima's D, Fu and Li's D, D*, F and F*, Fay and Wu's H and E, McDonald–Kreitman test, Fixation Index, and Extended Haplotype Homozygosity.
Our tool, CATE (CUDA Accelerated Testing of Evolution) is built upon two main features: a file organization system coupled with a novel search engine; and the large-scale parallelization of the algorithms using the GPU. First, a single file can contain well over a few million polymorphisms spanning thousands of individuals making them time-consuming to read and process sequentially. Therefore, we have split these VCF files based on the number of SNPs and positions. This enables random accessing of the genomic data for which we have designed an algorithm dubbed "compound interpolation search". It is a synergistic implementation of the interpolated search algorithm coupled with a multithreaded form of sequential search. This has greatly helped increase the speed with which information is collected while minimizing the demand for RAM. Second, we developed bespoke algorithms for each test. These algorithms harness the full potential of the CUDA core architecture to better process the polymorphic data and conduct the required computations.
[RESULTS]
With these implementations, our tool processes about 5200 genes per chromosome across all five super populations present in the 1000 Genome project in less than thirty minutes with an average time of six minutes per population. It is magnitudes faster than running standard tools such as vcf-kit and PopGenome.
GitHub repository: theLongLab/CATE: A fast and scalable CUDA implementation to conduct highly parallelized evolutionary tests on large scale genomic data. (github.com)

Quantifying the contribution of peer review to scientific publishing

Yashika Handa[1], Huaien Gu, Quan Long

1 University of Calgary, Calgary, AB
Presenting author: huaien.gu@ucalgary.ca

KEYWORDS
Peer review, natural language processing, web crawler, information retrieval, regression analysis

ABSTRACT
[Background] Although looks a bit exaggerated, the aphorism "Publish or Perish" somehow reflects a broadly appreciated pressure faced by scientific researchers. To succeed or simply survive in their academic career, researchers must publish, preferably in high-impact journals. Over the course, one needs to be evaluated by reviewers, who may propose comments that dramatically change the content of the work. Although the process of peer review normally improves the quality of the work, sometimes it also introduces substantial work that may not align to the authors' original plans. As a result, the final publication after peer review may be different to the authors' original intension. This might negatively impact authors' academic freedom. Despite all the scientists encounter this problem, its degree may vary with respects to different disciplines, funding systems, career stages of the authors, and different journals. Additionally, the contribution of peer review might not be welcomed by authors. These two questions motivate this project.

[Aims] We will use machine learning techniques to retrieve information from published papers to quantify the "difference" between the authors' original view and the post-peer review publication. We will also collect the features of authors and journals mentioned above. Using data collected, we will conduct standard statistical analysis to quantify associations between changes caused by peer review and the features of author/journals. Moreover, we will estimate the authors sentiment to the changes.

[Methods] The data collection pipeline is based on various web crawlers in conjunction with information retrieval (IR) techniques. We will start with the preprint servers including Arxiv [1], representing mathematics, statistics, physics, and computer science. Biorxiv [2], representing biological fields, and Medrxiv [3], representing medical research. The version #1 copies of these preprints are deemed as the "raw" version of the scientific works. Their subsequent versions (if exist) and formal publications after peer review will be deemed the refined work after peer reviews. The differences between them in terms of straightforward properties such as length, number of figure panels, number of authors, as well as complicated properties such as the scientific messages will be learned using IR techniques. The authors' view to the changes caused by reviewers will be quantified by their future self-citations and the quality new paragraphs. Finally, regression analysis and sentiment analysis will be conducted to quantify the statistical associations.

[Impacts] The deliverables of the work will bring quantify the contribution of peer review to academic independence and its variation in w.r.t. different features. As a by-product, this work will also reveal the difference between journals regarding to their editorial tradition in retaining authors' original work.

REFERENCES

1. https://arxiv.org
2. https://www.biorxiv.org
3. https://www.medrxiv.org

Transfer-learning enables accurate variant prioritization and improves power in transcriptome-wide association studies: an application to breast cancer

Qing Li[1], Xingyi Guo[2], and Quan Long[1]

1 University of Calgary, Calgary, Canada
2 Vanderbilt University School of Medicine, Nashville, USA

Presenting author: qing.li2@ucalgary.ca

KEYWORDS
Functional variants, transfer learning, gene expression, risk genes, transcriptome-wide association study

ABSTRACT
[Background] Empowered by huge samples and modern computing facilities (e.g., GPU and TPU), Machine Learning has been successful in resource-rich domains. However, some institutions may not enjoy the availability of large samples. Additionally, many researchers may not have state-of-the-art computational resources such as TPU to conduct deep training that is routinely happening in flagship companies. Transfer Learning provides a route to re-task a complicated model that has been pre-trained on a huge number of samples towards a different goal by limited rounds of training using small samples. An intuitive analog of transfer-learning is after observing cats and dogs, children can recognize wild animals like tigers and giraffes at first glance. Although children had never seen these animals before, they learned animals' features from previous observations of pets and transferred these experiences to complete new tasks—recognizing wild animals.
Enformer, a deep neural network trained by Google with high-performance TPU infrastructure, represents a state-of-the-art variant scorer[1]. It integrates 5,313 tracks of human multi-omics datasets as well as human and mouse reference genome patterns. Despite impressive functional information being integrated, it is not tailored to specific diseases, e.g., human cancers.
[Methods & Results] To re-task Enformer to human cancers and put more emphasis on transcription factors that have been shown to be critical by our recent publication[2], we have conducted a transfer learning by retraining the last three layers of Enformer on our in-house CHIP-Seq data for breast cancer. The outcome model, named "TF-Enformer" (Transcription Factors Enformer), can prioritize functional variants by integrating critical regulatory knowledge learned in Enformer and locally accessible transcription factors state tailored to breast cancer. It is shown that TF-Enformer is about 7% more accurate than Enformer when prioritizing genetic variants in breast-cell-related enhancers. To further demonstrate the use of TF-Enformer, we conducted a transcriptome-wide association study (TWAS) using these top variants weighted by their scores using the Breast Cancer Association Consortium data (N = 228,951). We have identified novel putative susceptibility genes including 19 unique genes unreported for

breast cancer. Further functional annotation showed our results outperform alternative models including the Enformer.

[Impacts] Our transfer learning framework will open the door to integrating the legacy model (Enformer) and dedicated knowledge and diseases. the newly identified genes will help decipher breast-cancer and its targeted therapy.

REFERENCES

[1] Z. Avsec *et al.*, "Effective gene expression prediction from sequence by integrating long-range interactions," (in English), *Nat Methods,* vol. 18, no. 10, pp. 1196-+, Oct 2021, doi: 10.1038/s41592-021-01252-x.

[2] W. Q. Wen *et al.*, "Genetic variations of DNA bindings of FOXA1 and co-factors in breast cancer susceptibility," (in English), *Nat Commun,* vol. 12, no. 1, Sep 13 2021, doi: ARTN 5318 10.1038/s41467-021-25670-9.

Effect of Antimicrobial Use in Natural Versus Conventional Cattle Feedlots on the Microbiome and Resistome

Catrione Lee [1,2], Rahat Zaheer [1], Krysty Munns [1], Gary Van Domselaar [3], Athanasios Zovoilis [2], Tim A. McAllister [1]

[1] University of Lethbridge, Lethbridge, Canada
[2] Agriculture and Agri-Food Canada, Lethbridge, Canada
[3] Public Health Agency of Canada, Winnipeg, Canada

Presenting author: catrione.lee@uleth.ca

KEYWORDS
Antimicrobial resistance, metagenomics, cattle production, next-generation sequencing, microbiome, resistome, *Methanobrevibacter*

ABSTRACT

Antimicrobial resistance (AMR) has been a rising concern for the clinical sector, agriculture, and the environment over the past 20 years. Tracking AMR using metagenomic approaches to identify antimicrobial resistant genes (ARGs) could facilitate the monitoring and assessment of risk across One Health continuums. In order to accomplish this goal, we have undertaken a study into the taxonomic and ARG profiles in water samples from feedlot catch basins (n=13) and cattle feces (n=60) from feedlots with differing management practices. Conventional (CONV) feedlots administered antimicrobials through injection and in feed to prevent and treat disease, while natural feedlots (NAT) do not administer antimicrobials. Samples were shotgun sequenced using Illumina HiSeq2000 (100-bp paired-end reads), classified by Kraken2, and analyzed by the AMR++ workflow. In comparing fecal samples, there was a 2.44- and 2.34-fold increase ($P < 0.001$) in the normalized read abundance of the genera *Methanobrevibacter* and *Treponema*, respectively, in CONV over NAT feedlots. Catch basin samples showed no differences ($P > 0.05$) in abundance of prevalent taxa between CONV and NAT feedlots. Using the MEGAres2.0 database in the AMR++ workflow identified no differences in either fecal or catch basin samples collected from CONV and NAT feedlots. However, all samples had high levels of ARGs encoding for tetracycline and macrolide resistance. Work to identify the relationship of microbial profiles and ARGs to genes coding for biocide and heavy metal resistance, along with their association with mobile genetic elements is ongoing. We hypothesize that metagenomics is posed to play a key role in the One Health risk assessment of AMR in livestock and the environment.

Genomic characterization of *Enterococcus hirae* from beef cattle feedlots and associated environmental continuum

Sani-e-Zehra Zaidi [1,2], Rahat Zaheer[1], Athanasios Zovoilis[2] & Tim A. McAllister[1]

1. Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, 5403 1st Avenue South, Lethbridge, AB, T1J 4P4, Canada.

2. Department of Chemistry and Biochemistry, University of Lethbridge, 4401 University Drive, Lethbridge, Alberta T1K 3M4

Presenting author: saniezehra.zaidi@uleth.ca

KEYWORDS

Cattle production, antimicrobial resistance, enterococci, comparative genomic analysis, genomic signatures

ABSTRACT

Enterococci are commensal bacteria of the gastrointestinal tract of humans, animals and insects. They are also found in soil, water and plant ecosystems. The presence of enterococci in human, animal and environmental settings make these bacteria ideal candidates to study antimicrobial resistance in the One-Health continuum. This study focused on *Enterococcus hirae* isolates (n= 4601) predominantly isolated from beef production systems including bovine feces (n=4117, 89.5%), catch-basin water (n=306, 66.5%), stockpiled bovine manure (n=24, 0.5%), and natural water source near feedlots (n=145, 32%), and few isolates from urban wastewater (n=9, 0.2%) denoted as human-associated environmental samples. Antimicrobial susceptibility profiling of a subset (n=1319) of *E. hirae* isolates originating from beef production systems (n=1308) showed high resistance to tetracycline (65%) and erythromycin (57%) with 50.4% isolates harboring multi-drug resistance, whereas urban wastewater isolates (n=9) were resistant to nitrofurantoin (44.5%) and tigecycline (44.5%) followed by linezolid (33.3%). Genes for tetracycline (*tetL, M, O , S/M and O/32/O*) and macrolide resistance *erm(B)* were frequently found in these beef production isolates. Antimicrobial resistance profiles of *E. hirae* isolates recovered from different environmental settings appeared to reflect the kind of antimicrobial usage in beef and human sectors. Comparative genomic analysis of *E. hirae* isolates showed an open pan-genome that consisted of 1427 core genes, 358 soft core genes, 1701 shell genes and 7969 cloud genes. Across species comparative genomic analysis conducted on *E. hirae, Enterococcus faecalis* and *Enterococcus faecium* genomes revealed that *E. hirae* had unique genes associated with vitamin production, cellulose and pectin degradation, traits which may support its adaptation to the bovine digestive tract. *E. faecium* and *E. faecalis* more frequently harbored virulence genes associated with biofilm formation, iron transport, and cell adhesion, suggesting niche specificity within these species.

cLD: Rare-variant disequilibrium between genomic regions identifies novel genomic interactions

Dinghao Wang[1,*], Jingni He[1,*], Deshan Perera[1,*], Chen Cao[1], Pathum Kossinna[1], Qing Li[1], William Zhang[2], Alexander Platt[3], Jingjing Wu[1], Qingrun Zhang[1]

[1] University of Calgary, Calgary, AB
[2] The Harker School, San Jose, CA, USA
[3] Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA USA
* joint first authors

Presenting author: dinghao.wang@ucalgary.ca

KEYWORDS
cumulative Linkage disequilibrium, asymptotic distribution, bootstrap, gene-gene interaction, case/control association studies

ABSTRACT
Linkage disequilibrium (LD) is a fundamental concept in genetics; critical for studying genetic associations
and molecular evolution. However, the calculation of LD involves the use of allele frequencies of the genetic variants in its denominator to normalize the statistic and therefore suffers from a high variance (instability) when allele frequencies are close to zero. As such, in practice, researchers only analyze common genetic variants with minor allele frequency (MAF) higher than a threshold (e.g.,0.05), excluding
more than 90% of human genetic variants.
In this work, we introduce cumulative LD (cLD), a stable statistic that captures the rare-variant LD between
genetic regions and opens the door for furthering biological knowledge using rare genetic variants. We also derive the asymptotic properties of cLD and LD using multinomial distributions and their multivariate
normal approximation as well as the multivariate Delta Method. Additionally, following the conventional statistical procedure of bootstrapping to empirically estimate stability, we sub-sample half of each population sample to form bootstrapped distributions for both cLD and LD. The subsampling shows that cLD exhibits a much slimmer bootstrapped distribution than LD across three ethnic groups.
In application, we find cLD reveals an increased genetic association between genes in 3D chromatin interactions, a phenomenon recently reported negatively by calculating standard LD between common variants. Additionally, we show that cLD is higher between gene pairs reported in interaction databases, identifies unreported protein-protein interactions, and reveals interacting genes distinguishing case/control samples in association studies.
The full details of this work may be found in our preprint (https://www.biorxiv.org/content/10.1101/2022.02.16.480745v1)

Processing of Alu and B2 SINE RNAs by Hsf1 is dysregulated in neuro-toxicity models and Alzheimer's disease patients

Luke Saville[1,2,3,4], Yubo Cheng[1,2,3,4], Babita Gollen[1,2,3,4], Chris Isaac[1,2,3,4], Jogender Mehla [4], Ana Alvarez Veronesi[5,6], Jeffrey T. Joseph[5,6], Majid Mohajerani[4], Athanasios Zovoilis[1,2,3,4]

Affiliations:
(1) Dept. of Chemistry and Biochemistry, University of Lethbridge, Canada
(2) Southern Alberta Genome Sciences Centre, University of Lethbridge, Canada
(3) Alberta RNA Research and Training Institute
(4) Canadian Center for Behavioural Neuroscience
(5) Departments of Pathology and Clinical Neurosciences, University of Calgary
(6) Calgary Brain Bank

Presenting author: luke.saville@uleth.ca

KEYWORDS

Alzheimer's disease, neural transcriptomics, ncRNAs, SINE RNAs, Hsf1

ABSTRACT

More than 98% of the mammalian genome is composed of non-coding regions. The non-coding SINE elements appear in millions of copies throughout the genome and are transcribed by RNA polymerase III (Pol-III) into SINE RNAs, such as the Alu and B2 RNAs of the human and mouse cells. Known to inhibit RNA polymerase II (Pol-II) at gene specific sites, Alu and B2 RNAs are potent regulators of large gene networks under dynamic cellular processes. Previously, the polycomb group protein, EZH2 was shown to process Alu and B2 RNAs, relieving Pol-II of its inhibition at stress response gene sites in response to heat shock. Similarly, when we examine Alzheimer's disease $APP^{+\ NL\text{-}G\text{-}F}$mouse model hippocampus tissue, we report aberrant processing of B2 RNA is correlated with the heat shock transcription factor, Hsf1, leading to cascading transcription changes of stress response genes (SRGs). Additionally, seeding the causative agent of neuro-toxicity in Alzheimer's disease (amyloid peptides) into a mouse hippocampal cell line, corroborates that B2 RNA processing and SRGs transcription are correlated, yet abrogated when Hsf1 is knocked down. Furthermore, we also report that human post-mortem brain tissue from Alzheimer's disease cohorts expresses aberrant processing patterns of Alu RNA in correlation with Hsf1 expression, TP53 expression, and stress response gene activation. With these findings, we present a body of work that examines the direct role of Hsf1 on Alu and B2 RNA processing and how Alu and B2 RNAs regulate stress response gene transcription in neuro-toxicity models and Alzheimer's disease.

An expression-directed linear mixed model (edLMM) discovering low-effect genetic variants

Jiayi Bian[1], Qing Li[2], Pathum Kossinna[2], Xiang Zhou[3], Jingjing Wu[1], Quan Long[1,2,4,5]

[1] University of Calgary, Department of Mathematics and Statistics, Calgary, Canada
[2] University of Calgary, Department of Biochemistry & Molecular Biology, Calgary, Canada
[3] University of Michigan, School of Public Health, Ann Arbor, USA
[4] University of Calgary, Hotchkiss Brain Institute, Calgary, Canada
[5] University of Calgary, Alberta Children's Hospital Research Institute, Calgary, Canada

Presenting author: jiayi.bian@ucalgary.ca

KEYWORDS
Low-effect genetic variants, Transcriptome-wide association studies (TWAS), Linear mixed models (LMMs), Genetically regulated expression (GReX), expression-directed linear mixed model (edLMM)

ABSTRACT
[Background] Discovering genetic variants associated with complex traits is a long-standing theme in genome-wide association studies (GWAS). Given complicated genetic background of participating individuals, it is generally difficult to identify genetic variants with small effect sizes with a moderate sample size. Although their individual contributions are small, aggregately these low-effect variants might account for a substantial proportion of the heritability.

[Rationale] The transcriptome-wide association studies (TWAS) is a pioneering approach utilizing gene expression data to identify genetic basis of complex diseases. Its core component, genetically regulated expression (GReX), links gene expression with phenotype by serving as both the outcome of genotype-based expression models and the predictor for downstream association testing [1,2]. Although it has been used in many high-profile projects, its mathematical interpretation hasn't been rigorously verified. As such, we first conducted power analysis using NCP-based closed forms and realized that the common interpretation of TWAS that looks biologically sensible is mathematically questionable [3]. Following this insight, by real data analysis and simulations, we demonstrated that current linear models of GReX inadvertently combine two separable steps of machine learning - feature selection and aggregation - which can be independently replaced to improve overall power [4]. Based on this new interpretation, we have developed novel protocols disentangling feature selections and aggregations, leading to improved power and novel biological discoveries [5].

[Methods] Linear mixed models (LMMs) serve as a primary approach to discovering genetic variants associated with disease phenotype in association studies [6]. This is largely due to the advantage of using a random term modeled by the genetic relationship matrix (GRM). In this work, based on our understanding of LMMs and alternative interpretation of TWAS, we developed an expression-directed linear mixed model (edLMM), a synergy between LMM and TWAS. By utilizing the functional weights learned from transcriptome data, we formed an alternative approach to estimate this random term, which can estimate the genetic background more accurately.

[Results] By applying edLMM to cohorts of around 5,000 individuals with phenotype of either binary (Wellcome Trust Case Control Consortium) or quantitative (Northern Finland Birth Cohort 1966) traits, we demonstrated edLMM reported more significant genetic variants and they are functionally more

relevant. In both real data and simulations, we showed edLMM is more powerful than EMMAX, a flagship LMM used by many researchers, and it is particularly sensible when the effect size is low. Additionally, the SNP-based heritability is substantially improved when the low-effect variants discovered by edLMM are incorporated.

REFERENCES

1. https://pubmed.ncbi.nlm.nih.gov/26258848/
2. https://pubmed.ncbi.nlm.nih.gov/34021117/
3. https://pubmed.ncbi.nlm.nih.gov/33635859/
4. https://pubmed.ncbi.nlm.nih.gov/33200776/
5. https://pubmed.ncbi.nlm.nih.gov/34849857/
6. https://pubmed.ncbi.nlm.nih.gov/22706312/

Identification of a genetically divergent population of the human soil transmitted helminth Trichuris in Cote d'Ivoire that is unresponsive to albendazole-ivermectin combination treatment

Abhinaya Venkatesan[1], Rebecca Chen[1], Brenna Reimer[1], Somphou Sayasone[3], Said Ali[4], Jean Coulibaly[5], Chandni Patel[2], Sophie Welsche[2], Ladina Keller[2], Eveline Hürlimann[2], Jennifer Keiser[2], John S Gilleard[1]

Affiliations:
1. Faculty of Veterinary Medicine, University of Calgary, Canada
2. Swiss Tropical and Public Health Institute, Basel, Switzerland
3. Lao Tropical and Public Health Institute, Vientaine, Lao PDR
4. Public Health Laboratory Ivo de Carneri, Zanzibar, Tanzania
5. Centre Suisse de Recherches Scientifiques en Côte d'Ivoire, Abidjan, Côte D'Ivoire

Presenting author: abhinaya.venkates1@ucalgary.ca

KEYWORDS
Put Keywords Here, separated by commas, should be at least 5 keywords
Deep amplicon sequencing, Trichuriasis, Drug efficacy, Genetic diversity, Amplicon Sequence Variants (ASVs)

ABSTRACT
Trichuris trichiura, an important intestinal parasitic nematode, infects over 500 million people globally and various studies have identified cryptic species of *Trichuris* infecting humans and non-human primates [1]. Treatment includes Mass Drug Administration (MDA) programs that use either albendazole or mebendazole, but these benzimidazole drugs have very low efficacy against *T. trichiura,* with egg reduction rates (ERR) typically below 50% [2]. However, albendazole-ivermectin combination treatment has considerably improved efficacy and is being considered by the World Health Organization (WHO) in wide-scale MDA programs. Recently, parallel albendazole-ivermectin clinical efficacy trials were performed in three regions – Laos, Tanzania, and Cote d'Ivoire, and revealed high efficacy in Tanzania and Laos (ERR >98%) but much lower efficacy in Cote d'Ivoire (ERR <70%) [3]. To explore whether this difference in efficacy could be due to genetic differences in the parasite populations and/or the presence of cryptic species, we performed deep amplicon sequencing of multiple mitochondrial and ribosomal DNA taxonomic markers on *T. trichiura* PCR positive fecal samples from the three regions. Primers were designed to target the ribosomal ITS-1 and ITS-2, mitochondrial *nad1*, *nad4, cox-1,* and the major β-tubulin gene and the PCR amplicons were sequenced at depth from stool DNA samples on the Illumina MiSeq platform. Analysis of the mitochondrial and β-tubulin sequences generated Amplicon Sequence Variants (ASVs) mapping to the appropriate reference sequences from all the samples from Tanzania and Laos, but not from Cote d'Ivoire. Phylogenetic analysis of the ribosomal ITS-1 and ITS-2 loci revealed much lower ASV diversity and extreme genetic divergence in the Cote d'Ivoire population compared to Tanzania and Laos. Phylogenetic and haplotype network analysis revealed that the divergent *Trichuris* sequences were more closely related to *Trichuris* from non-human primates and *Trichuris suis* than *T. trichiura*. In summary, this study has identified a genetically divergent *Trichuris* population in Cote d'Ivoire, which is likely to be a cryptic species, and has a significantly lower sensitivity to albendazole-ivermectin combination treatment than *T. trichiura* populations. Further work is required to explore the distribution and the potential of this cryptic *Trichuris* species to compromise the effectiveness of albendazole-ivermectin MDA programs introduced by the WHO.

REFERENCES

1. *https://pubmed.ncbi.nlm.nih.gov/25340752/*
2. *https://doi.org/10.1136/bmj.j4307*
3. *https://pubmed.ncbi.nlm.nih.gov/34856181/*

Developing a modular resource to support comprehensive and flexible long-read and short-read rDNA metabarcoding across the nematode phylum

Eléonore Charrier[1], Rebecca. Chen[1], John. S. Gilleard[1]

[1] Faculty of Veterinary Medicine, Host-Parasite Interactions Program (HPI), University of Calgary, Calgary, Canada

Presenting author: Eléonore Charrier, elonore.charrier@ucalgary.ca, Faculty of Veterinary Medicine, Host-Parasite Interactions Program (HPI), University of Calgary, Calgary, Canada

KEYWORDS
Metabarcoding, ITS-2, Database, 18S rDNA, rDNA cistron, Nematodes

ABSTRACT

Nematodes are important human, animal and plant pathogens, as well as important free-living organisms, that are underexplored at the community level relative to many other groups. ITS-2-rDNA metabarcoding using short read Illumina sequencing was established to investigate strongylid nematode communities in domestic livestock for diagnostic and epidemiological studies [1,2]. However, the ITS-2 rDNA marker shows limitations as a marker with sequence diversity either being too low to distinguish some very closely related species or too high to resolve deeper phylogenetic relationships. This is particularly challenging for poorly defined complex parasitic nematode communities such as wildlife, or free-living nematode communities. Consequently, we are developing a modular approach using different taxonomic rDNA markers including the 18S and 28S rDNA coding regions to resolve the deeper phylogenetic relationships and the combined ITS-1-5.8S-ITS-2 region to resolve more shallow phylogenetic relationships. The rDNA cistron reference sequences present in public databases are redundant, often partial, and sometimes misannotated, and so accurate, well-curated, and regularly updated bespoke databases are needed to support accurate species assignment from metabarcoding data.

The markerDB pipeline was previously developed and used to produce a nematode ITS2 rDNA database supporting ITS-2 rDNA short-read nemabiome metabarcoding [5]. We have produced a set of similar databases for different regions of the rDNA cistron (18S, 28S, ITS-1 and ITS-1,5.8S,ITS-2 regions). We are also assessing the "phylogenetic coverage" of both established (eg. NC1, NC2, NC5, NC13) and new primer targeting the nematode 18S, 28S and ITS-1,5.8S,ITS-2 rDNA regions [3,4].

Manual searches of the NCBI Genbank database were first undertaken to determine the search terms and sequence length ranges to obtain the maximum number of sequences for each target. These criteria where then used to parameterize our markerDB pipeline to search Genbank and recovered sequences were filtered based on the identify of upstream and downstream flanking regions to select the full-length reference sequences. The non-redundant full-length databases currently comprise 2570, 227, 11849 and 9181 unique sequences representing 1381, 196, 1708 and 1248 nematode species respectively. We assessed 25 primers targeting the 18S rDNA and 19 primers targeting the 28S rDNA coding regions, as well as 2 primers targeting the 5.8S rDNA coding region. Data visualisations showing the phylogenetic coverage of primer sites will be presented .

This modular set of rDNA marker databases, and primer sets with defined phylogenetic coverage, will support more flexible, customized, and comprehensive long-read and short-read metabarcoding of both parasitic and free-living nematode communities.

REFERENCES

1. Avramenko RW, Redman EM, Windeyer C, Gilleard JS (2020). Assessing anthelmintic resistance risk in the post- genomic era: a proof-of-concept study assessing the potential for widespread benzimidazole-resistant gastrointestinal nematodes in North American cattle and bison. Parasitology 147, 897–906. https:// doi.org/10.1017/S0031182020000426
2. Queiroz, C., Levy, M., Avramenko, R., Redman, E., Kearns, K., Swain, L., Silas, H., Uehlinger, F., & Gilleard, J. S. (2020). The use of ITS-2 rDNA nemabiome metabarcoding to enhance anthelmintic resistance diagnosis and surveillance of ovine gastrointestinal nematodes. *International Journal for Parasitology: Drugs and Drug Resistance*, *14*(June), 105–117. https://doi.org/10.1016/j.ijpddr.2020.09.003
3. Newton, L. A., Chilton, N. B., Beveridge, I., Hoste, H., Nansen, P., & Gasser, R. B. (1998). Genetic markers for strongylid nematodes of livestock defined by PCR-based restriction analysis of spacer rDNA. *Acta Tropica*, *69*(1), 1–15. https://doi.org/10.1016/S0001-706X(97)00105-8
4. Gasser, R. B., Chilton, N. B., Hoste, H., & Beveridge, I. (1993). Rapid sequencing of rDNA from single worms and eggs of parasitic helminths. *Nucleic Acids Research*, *21*(10), 2525–2526. https://doi.org/10.1093/nar/21.10.2525
5. Workentine, M. L., Chen, R., Zhu, S., Gavriliuc, S., Shaw, N., Rijke, J. de, Redman, E. M., Avarmenko, R. W., Wit, J., Poissant, J., & Gilleard, J. S. (2019). A database for ITS2 sequences from nematodes. *BioRxiv*, 689695. https://doi.org/10.1101/689695

A transcriptomic view of the molecular mechanisms that the lancet liver fluke uses to manipulate its hosts behavior

Chenhua Li[1], Cameron P. Goater[2], James D. Wasmuth[1]

1 University of Calgary, Calgary, Alberta, Canada
2 University of Lethbridge, Lethbridge, Alberta, Canada

Presenting author: chenhua.li1@ucalgary.ca

KEYWORDS
Host manipulation, parasites, behavior, transcriptome, RNA-sequencing

ABSTRACT
The lancet liver fluke, *Dicrocoelium dendriticum*, is an invasive species, recently discovered in Alberta. The parasite's life-cycle is complex, passing through snails, ants and ruminants, e.g. deer, sheep and cattle. Each host is manipulated by *D. dendriticum* to promote the parasite's survival and transmission. Larvae of *D. dendriticum* manipulate infected ants to leave their nests during the cool hours of the day, climb to the top of a flower, where they attach themselves to a petal with their mandibles. These ants remain attached overnight, detach the next morning when temperatures rise, and return to their nest. They repeat this bizarre attach/detach sequence for the rest of the summer. Uninfected ants from the same nest do not engage in these odd behaviors. We seek to understand the molecular mechanisms that underlie this complex manipulation of host behavior. Using transcriptomics, we compared the gene expression pattern of brains from infected and uninfected *Formica aserva* collected from a site of *D. dendriticum* emergence in southern Alberta, Canada. We recreated the manipulation cycle in the lab and measured RNA from ant brains at four different stages of the behavior manipulation, generating RNA-Seq data for each. We found 3295 genes that were differentially expressed between infected and uninfected ants. The functions of these genes included environmental sensing (odorant, vision, gustatory), circadian rhythm, immune response, muscle contraction, the production of biogenic monoamines, and certain hormones. Genes involved in odorant and vision were downregulated in attached infected ants. Vision genes were upregulated in post-attached infected ants compared to uninfected controls. In infected ants, muscle genes were upregulated during the pre-attached stage and then downregulated at the post-attached stage. Genes involved in serotonin synthesis were also downregulated during the post-attached stages in infected ants, implicating biogenic monoamines as key regulators of this behavior change. With these results, we better understand how *D. dendriticum* manipulates their ant host's behavior.

Microbiome analysis of tailings reclamation protocols in the Athabasca Oil Sands Region

Elisabeth Richardson[1], Muhammad Arslan[2], Mohamed Gamal El-Din[2], and Patrick Hanington[1]

[1] School of Public Health, University of Alberta, Edmonton, Canada
[2] Department of Civil and Environmental Engineering, Faculty of Engineering, University of Alberta, Edmonton, Canada

Presenting author: ehrichar@ualberta.ca

KEYWORDS
microbiome, reclamation, oil sands, amplicon sequencing, environmental DNA

ABSTRACT

The Athabasca Oil Sands Region is a region of Northern Alberta that contributes enormously to the province's heritage, environment, and economy. The process of bitumen extraction from oil sands results in the production of liquid mining waste, known as tailings, and methods for efficient reclamation of this waste is a research priority for Alberta. At Lake Miwasin, a demonstration pit lake on the Suncor mine site, PASS technology (Permanent Aquatic Storage Structure) is being used to accelerate dewatering of tailings into a more stable state amenable to further reclamation. Concurrently, mesocosms of PASS-treated tailings capped with oil sands process water were established in the Department of Engineering at the University of Alberta in 2021.

We present an environmental DNA (eDNA) -based analysis of the microbiome of these mesocosms in the first year of monitoring. We discuss some of the challenges associated with extraction of eDNA for sequencing from hydrocarbon-associated environment, and the methods we have developed to ensure consistent, high-quality DNA yields. We have identified distinct microbial communities associated with the sediment layer, consisting of PASS-treated tailings, and overwater. Both communities appear to be most significantly associated with oxygen availability, and the sediment layer is dominated by anoxic taxa. Altogether, these results indicate a diverse and thriving bacterial community within both the PASS-treated tailings and the overwater cap, the composition of which can inform the adaptive management strategy of the demonstration pit lake Lake Miwasin during the reclamation process.

Rejuvenation Effects of Phytocannabinoids and Anti-aging Drugs on Dermal Fibroblasts During Aging

Marta Gerasymchuk[1], Gregory Robinson[1], Olga, Kovalchuk[1], Igor Kovalchuk[1]
[1] University of Lethbridge, Lethbridge, AB, T1K3M4, Canada

Presenting author: marta.gerasymchuk@uleth.ca

KEYWORDS
Aging, Dermal Fibroblasts, Phytocannabinoids, Anti-aging Drugs, Rejuvenation

ABSTRACT
Identifying effective anti-aging compounds is a cornerstone of modern longevity, aging and skin health research. There are numerous experimental data on the ability of certain drugs and natural compounds to prevent, delay, or alleviate the development of age-related diseases, their symptoms, cellular senescence, and rejuvenation activity. *Cannabis sativa L*. and the numerous phytocannabinoids (pCBs) found within, including delta-9-tetrahydrocannabinol (THC) and cannabidiol (CBD), have been known for millennia to have many therapeutic uses. Cannabinoids are effective in numerous pathological conditions such as fibrosis and skin disorders, however, there is a lack of scientific knowledge about the anti-aging and rejuvenation properties of cannabinoids. There is considerable evidence of the effectiveness of natural signaling regulators (NSRs) such as metformin, triacetylresveratrol (TRSV), and rapamycin in longevity and anti-aging studies. Concomitantly, there is controversial information regarding their potential protective role against skin aging. In addition, no information exists about the combined effects of NSRs and pCBs on dermal aging and potential rejuvenation properties.
Here we tested the efficacy of metformin, TRSV, and rapamycin combined with pCBs as anti-aging compounds on the CCD-1064Sk, CCD-1135Sk, and BJ-5ta skin fibroblasts cell lines. The nuclear alterations based on the DAPI staining were analyzed. The cellular viability was determined using the MTT assay. The ability of NRS combined with pCBs to potentiate the functional and metabolic activity of the senescent fibroblasts was assessed based on the RT-PCR data and wound healing assay. We found that metformin and TRSV combined with pCBs reduced the adverse influence of oxidative stress, lowered expression of senescence markers, and in turn, stimulated metabolic and functional activity, inhibited dermal aging processes, and positively affected the viability of skin fibroblasts. Therefore, pCBs can be a valuable source of biologically active substances used in aging. Concurrently, the data obtained from three different dermal fibroblasts cell lines demonstrated rapamycin's adverse effects on the speed of wound healing in senescent cultures. The analysis also showed that cannabinoids alone and combined with TRSV positively affected the wound healing process. Thus, TRSV alone or combined with pCBs have a potential for enhancing regeneration and repair in injured tissues. More studies are needed, especially those using 3D tissues and animals, to confirm the efficacy of pCBs combined with metformin or TRSV on various processes associated with aging.

Identification of mimicry between protein structures and its role in host-parasite interactions

Viraj Muthye[1], Kaylee D. Rich[1], James D. Wasmuth[1]

[1] University of Calgary, Calgary, Alberta, Canada

Presenting author: virajrajendra.muthye@ucalgary.ca

KEYWORDS
Molecular mimicry, malaria, *Plasmodium*, protein structure, host-parasite interactions

ABSTRACT

Parasitic organisms use a dazzling array of molecular mechanisms to manipulate their host. One mechanism, termed "molecular mimicry", is a parasite protein evolving to share structural and/or sequence similarity with host proteins co-opting their function. For example, in malaria, the apical membrane antigen (AMA-1) helps the parasite, *Plasmodium falciparum*, attach to the host's red blood cell. AMA-1 is structurally similar to the PAN protein domain in plasminogen, a protein that mediates protein-carbohydrate interactions at the surface of red blood cells. Large-scale searches for structural similarity have been impractical until very recently. The search for molecular mimicry candidates has relied on finding unexpected sequence similarities between parasite proteins and host proteins. However, we show that widely-used sequence-based methods have multiple shortcomings and can result in an unacceptably high noise-to-signal ratio. To improve on this, we created a pipeline that searches for unexpected 3D structural similarities between parasite proteins and host proteins. We used protein structures for *P. falciparum*, human, and 15 control species (species not infected by *P. falciparum*) from the Protein Data Bank (PDB) and generated by AlphaFold2. We aligned these structures with Foldseek and found 72 *Plasmodium* proteins that aligned to 69 human proteins with a better score than proteins from the control species. However, only a few known *Plasmodium*-human interacting proteins were on this list. Further, gene ontology analysis did not identify expected functions statistically enriched in the mimicked human proteins, e.g., immune system and cell adhesion. From these data, we conclude that in malaria at least, molecular mimicry is not a major evolutionary mechanism for host-parasite interactions.

Whole-genome sequencing analysis of clozapine-induced myocarditis

Ankita Narang*[1], Paul Lacaze[2], Kathlyn J Ronaldson[2], John J McNeil[2], Mahesh Jayaram[3,4], Naveen Thomas[3,4], Rory Sellmer[5], David Crockford[5], Robert Stowe[6], Steven C. Greenway[7,8,9,10,11], Christos Pantelis[3,4], Chad A Bousman[1,5,9,12,13]

[1]Department of Medical Genetics, University of Calgary, Calgary, AB, Canada.

[2]Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, VIC, Australia

[3]Adult Mental Health Rehabilitation Unit, North Western Mental Health, Melbourne Health, Western Health, Sunshine Hospital, St Albans, VIC, Australia

[4]Melbourne Neuropsychiatry Centre, Department of Psychiatry, The University of Melbourne and Melbourne Health, Western Centre for Health & Education, St Albans, VIC, Australia

[5]Department of Psychiatry, University of Calgary, Calgary, AB, Canada

[6] Departments of Psychiatry and Neurology (Medicine) and Neuropsychiatry Program, University of British Columbia, Vancouver, BC, Canada

[7]Department of Pediatrics, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[8]Department of Biochemistry & Molecular Biology, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[9]Alberta Children's Hospital Research Institute, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[10]Department of Cardiac Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[11]Libin Cardiovascular Institute, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[12]Department of Physiology & Pharmacology, University of Calgary, Calgary, AB, Canada.

[13]Hotchkiss Brain Institute, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Presenting author: ankita@ucalgary.ca

ABSTRACT

Introduction: One of the concerns limiting the use of clozapine in schizophrenia treatment is the risk of rare but potentially fatal myocarditis [1]. Our previous genome-wide association study and human leucocyte antigen analyses identified putative loci associated with clozapine-induced myocarditis [2]. To build on these promising findings, we undertook the first whole-genome sequencing study to examine different classes of DNA variation not measured in our previous genetic study, including rare deleterious coding variants, copy number variants, and candidate variation in pharmacogenes associated with clozapine metabolism.

Methods: Whole-genome sequencing was performed on 25 cases with clozapine-induced myocarditis and 25 demographically-matched clozapine-tolerant control subjects. *The Illumina DRAGEN* (version

07.021.382.3.4.9, Dynamic Read Analysis for GENomics) Bio-IT platform was used for alignment and variant calling. Stargazer v1.08 [3] tool was used to assign genotype-inferred metabolizer phenotypes for cytochrome P450 genes. We performed burden analysis using rare and deleterious variants, defined based on the minor allele frequency from global databases and combined annotation dependent depletion (CADD) annotation score using ANNOVAR software [4].

Results: We identified 15 genes based on rare variant gene-burden analysis (*MLLT6, CADPS, TACC2, L3MBTL4, NPY, SLC25A21, PARVB, GPR179, ACAD9, NOL8, C5orf33, FAM127A, AFDN, SLC6A11, PXDN*) nominally associated ($p<0.05$) with clozapine-induced myocarditis. Of these genes, 13 were expressed in human myocardial tissue.

Conclusion: Our study provides preliminary insights into the potential role of rare genetic variants in susceptibility to clozapine-induced myocarditis. Although these candidate genes will require independent validation and their roles in the development of clozapine-induced myocarditis remain to be tested.

REFERENCES

1. https://pubmed.ncbi.nlm.nih.gov/10584719/
2. https://pubmed.ncbi.nlm.nih.gov/32066683/
3. https://pubmed.ncbi.nlm.nih.gov/31206625/
4. https://pubmed.ncbi.nlm.nih.gov/20601685/

Type B Ultra Long-Range Interactions in PFAs (TULIPs) represent recurrent epigenomic alterations in pediatric ependymoma

Michael J Johnston[1,2]*, John JY Lee[3,4,5]*, Bo Hu[6,7]*, Ana Nikolic[1,2,8]*, Audrey Baguette[9], Seungil Paik[1,2], Haifen Chen[6], Sachin Kumar[3,4,5], Carol CL Chen[6], Selin Jessa[9], Polina Balin[3,4,5], Vernon Fong[3,4,5], Melissa Zwaig[6], Kulandaimanuvel A Michaelraj[3,4], Xun Chen[7], Yanlin Zhang[10], Srinidhi Varadharajan[11], Pierre Billon[1,8], Nikoleta Juretic[12], Craig Daniels[3,4], Amulya Nageswara Rao[13], Caterina Giannini[13], Eric M Thompson[14], Miklos Garami[15], Peter Hauser[15], Timea Pocza[15], Young Shin Ra[16], Byung-Kyu Cho[17], Seung-Ki Kim[17], Kyu-Chang Wang[17], Ji Yeoun Lee[17], Wieslawa Grajkowska[18], Marta Perek-Polnik[19], Sameer Agnihotri[20], Stephen Mack[11], Benjamin Ellezam[21], Alex Weil[22], Guillaume Bourque[6,23], Jennifer A Chan[1,2,24], Mathieu Lupien[25,26,27], Jiannis Ragoussis[6,23], Claudia Kleinman[6,28], Jacek Majewski[6], Nada Jabado[6,12,29§], Michael D Taylor[3,4,5§], Mathieu Blanchette[9§], Marco Gallo[1,2,8§]

[1]Arnie Charbonneau Cancer Institute, Cumming School of Medicine, University of Calgary, Calgary, AB T2N 4N1, Canada
[2]Alberta Children's Hospital Research Institute, Cumming School of Medicine, University of Calgary, Calgary, AB T2N 4N1, Canada
[3]The Arthur and Sonia Labatt Brain Tumor Research Center, The Hospital for Sick Children, Toronto, ON M5G 0A4, Canada
[4]Developmental & Stem Cell Biology Program, The Hospital for Sick Children, Toronto, ON M5G 0A4, Canada
[5]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON M5G 1L7, Canada
[6]Department of Human Genetics, McGill University, Montreal, QC H2A 1B1, Canada
[7]Department of Anatomy and Cell Biology, Kyoto University, Kyoto 606-8501, Japan
[8]Department of Biochemistry and Molecular Biology, Cumming School of Medicine, University of Calgary, Calgary, AB T2N 4N1, Canada
[9]Quantitative Life Sciences, McGill University, Montreal, QC H3A 1B9, Canada
[10]School of Computer Science, McGill University, Montreal, QC H3A 2A7, Canada
[11]Department of Developmental Neurobiology, St Jude Children's Research Hospital, Memphis, TN 38105, United States of America
[12]Department of Pediatrics, McGill University, and The Research Institute of the McGill University Health Centre, Montreal, QC H4A 3J1, Canada
[13]Pediatric Hematology-Oncology, Mayo Clinic, Rochester, MN 55905, United States of America
[14]Department of Neurosurgery, Duke University Medical Center, Durham, NC 27710, United States of America
[15]Department of Pediatrics, Semmelweis University, Budapest H-1094, Hungary
[16]Department of Neurosurgery, University of Ulsan, Asan Medical Center, Seoul 05505, South Korea
[17]Department of Neurosurgery, Division of Pediatric Neurosurgery, Seoul National University Children's Hospital, Seoul 30322, South Korea
[18]Department of Pathology, The Children's Memorial Health Institute, University of Warsaw, Warsaw 04-730, Poland
[19]Department of Oncology, The Children's Memorial Health Institute, University of Warsaw, Warsaw 04-730, Poland
[20]Department of Neurosurgery, University of Pittsburgh Medical Center, Pittsburgh, PA 15213, United States of America
[21]Department of Pathology, Centre Hospitalier Universitaire Sainte-Justine, Université de Montréal, Montreal, QC H3T 1C5, Canada

[22]Department of Pediatric Neurosurgery, Centre Hospitalier Universitaire Sainte-Justine, Université de Montréal, Montreal, QC H3T 1C5, Canada
[23]McGill Genome Centre, Montreal, QC H3A 0G1, Canada
[24]Department of Oncology, Cumming School of Medicine, University of Calgary, Calgary, AB T2N 4N1, Canada
[25]Princess Margaret Cancer Centre, University Health Network, Toronto, ON M5G 1L7, Canada
[26]Department of Medical Biophysics, University of Toronto, Toronto, ON M5G 1L7, Canada
[27]Ontario Institute for Cancer Research, Toronto, ON M5G 0A3, Canada
[28]Lady Davis Research Institute, Jewish General Hospital, Montreal, QC H3T 1E2, Canada
[29]Division of Experimental Medicine, Department of Medicine, McGill University, Montreal, QC H3A 3J1, Canada

*Co-first authors
[§]Senior authors

Presenting author: Michael.Johnston1@ucalgary.ca

KEYWORDS
Brain tumors, pediatric cancer, Hi-C sequencing, multi-omics, epigenetics, chromatin

ABSTRACT

Posterior Fossa Group A (PFA) ependymomas are pediatric brain tumors with extremely poor survival outcomes. As protein-coding mutations in PFA are exceedingly rare [1], the underlying etiology of these tumors remains elusive. Elevated CpG island methylation [1] and depletion of H3K27me3 [2] have been described in PFA, leading to the hypothesis that PFA may be driven by a dysregulated epigenetic state. In this study, we sought to determine how three-dimensional (3D) genome features (such as DNA loops, domains, and compartments) differ between pediatric brain tumors. We performed Hi-C sequencing on a collection of 64 patient specimens and patient-derived primary cultures that collectively span multiple subgroups of ependymoma, medulloblastoma, high-grade glioma, and non-neoplastic brain. For certain samples, we further performed RNA-seq, histone modification ChIP-seq, or whole-genome bisulfite sequencing to allow multiomic data integration. Overall, the 3D genome organization of PFA samples appeared distinct from other tumor types. Small-scale looping interactions were reduced, while large, compartment-scale interactions were strengthened. We identified and defined TULIPs: a subset of type B compartments, separated by genomic distances >10 Mbp, that exhibit a striking >5-fold increase in reciprocal interaction strength. These TULIPs recurred at the same genomic positions across the vast majority of PFA samples with minimal representation among other tumor or non-tumor samples. TULIPs displayed enrichment for heterochromatic features such as H3K9me3 and late replication timing and were depleted of euchromatic features such as H3K27ac and protein-coding genes. Application of 3D chromosome modeling based on Hi-C data predicted aggregation of TULIP loci within the nucleus. By using immuno-fluorescence for H3K9me3 and oligo-FISH to label TULIP regions, we demonstrated that TULIP regions are more compact in PFA than other tumors. Finally, by applying inhibitors of H3K9 lysine methylation to PFA cultures we showed that TULIPs become more diffuse and cell viability is reduced. Altogether, this work defines TULIPs as highly recurrent epigenetic features of PFA tumors.

REFERENCES

1. https://doi.org/10.1038/nature13108
2. https://doi.org/10.1126/scitranslmed.aah6904

Site of Antibody Affinity Maturation in Early Vertebrates

Emmanuel Alimo, Bradley Magor.

University of Alberta, Edmonton Alberta, Canada

Presenting author: alimo@ualberta.ca

KEYWORDS

Antibody affinity maturation (AAM), Melanomacrophage cluster (MMc), Germinal center, autoantibodies, Zebrafish, Somatic Hypermutation.

ABSTRACT

For a long time, the process of Antibody Affinity Maturation (AAM) has been poorly understood in humans with autoimmune diseases. However, recent studies indicate that sites of affinity maturation in autoimmune individuals occurs in sites outside of the known germinal centers and these areas where the autoantibodies produced by B-cells affinity mature may be analogous to Melanomacrophage Clusters (MMCs) in early vertebrates. This study will provide more insight into where these MMCs may be located in the gut of early vertebrates, like Zebrafish. The fishes will be categorized into two groups as unvaccinated and vaccinated. Fluorescence microscopy will be used to detect and isolate the MMCs from the gut which autofluorescence due to the presence of melanin, hemosiderin and lipofuscin. B-cells expressing IgM and IgZ which are prominent in the gut will be targeted with specific primers and then amplified with PCR. Following the confirmation of the transcript size with gel electrophoresis, it will then be sequenced through next generational sequencing. This will give us an idea if the B-cells within the MMCs are undergoing Somatic Hypermutation which is a classical indication of B-cells undergoing antibody affinity maturation in germinal centers and this can be traced back to the parental cells that are have antigen bound to them before entering the site. Further, single cell RNA-seq will be done to see what other cell types (and their proportions) are involved in the antibody affinity maturation process. This work will not only have implications on vaccine use in aquaculture, but also help us better understand antibody affinity maturation processes in autoimmune disorders.

PacBio Sequencing to Explore Microbiome Communities of Aquaculture Environments

(Christine) Xiaoji Liu[1], Shawn M.C. Robinson[2]
[1] Agriculture and Agri-Food Canada, Lacombe, A.B., Canada
[2] Fisheries and Oceans Canada, St. Andrews, N.B., Canada

Presenting author: christine.liu2@agr.gc.ca

KEYWORDS
Microbiome, PacBio, Metagenomics, 16S, MicrobiomeAnalyst, Biocluster, Aquaculture, QIIME2, GitHub

ABSTRACT
Recent developments in sequencing technology, including PacBio, improves the resolution of microbiome communities to species/strain level by offering long read lengths without sacrificing the accuracy. Our study explored whether PacBio sequencing would be a suitable method to study microbiome communities of the aquaculture environment.
We sequenced the full-length 16S rRNA of 80 microbiome communities, including finfish (salmon skin and gut), benthic animals (crabs, lobsters, sea urchins, sea cucumbers, scallops), and sediment (hard and soft bottoms), on the PacBio Sequel II system at University of Illinois. The raw data was processed according to Microbiome Helper protocol [1] and QIIME2 [2]. The SILVA138 was used as the taxonomy reference database [3]. Rarefaction was performed on samples with a library size ≥ 500 without scaling. Taxa abundance, alpha- and beta- diversities were calculated through MicrobiomeAnalyst [4]. An average 23,192 reads per sample was obtained with the mean read length of 1,525 bps. The alpha-diversity analyses showed that sediment had significantly higher ($P<0.001$) diversity than other samples such as salmon or benthic animals. The microbiome communities of the skin mucus of farmed salmon (dominated by *Pseudoalteromonas/Vibrio*) was distinct from sediment (dominated by chloroplasts and halophilic sulfate-reducing bacteria) and salmon gut (dominated by Firmicutes/Bacterioidota). For benthic animals, lobsters were dominated by Vibrionaceae (68% total abundance), scallops dominated by Endozoicomonadaceae (63%), while crabs were dominated by Mycoplasmataceae (93%). Sea cucumbers were dominated by 36% Entomoplasmatales, 22% Chloroplast and 16% Bacillaceae; and sea urchins were dominated by 18% Fusibacteraceae, 17% Flavobacteriaceae,and 12% Desulfocapsaceae. Principal Coordinates Analyses further showed that resolution is best at the genus- and species-level for differentiating the microbiome communities between mucus of farmed salmon, sediment and salmon gut.
In conclusion, PacBio sequencing of the full-length 16S rRNA was able to characterize the unique fingerprints of microbiome communities at the genus-to-species level and to differentiate sample types in aquaculture environments.

REFERENCES

1. https://github.com/LangilleLab/microbiome_helper/wiki/PacBio-CCS-Amplicon-SOP-v1-(qiime2)
2. https://qiime2.org/
3. https://github.com/qiime2/docs/blob/master/source/data-resources.rst
4. https://www.microbiomeanalyst.ca/

Comparison of species taxonomy metabarcoding pipelines for bovine *Eimeria* COI and correlation with morphological species identification

Venkateswara Rao Parimisetti[1], Elizabeth Redman[1], Manigandan Lejeune[2], Holly White[2], Berit Bangoura[3], Nicolette Shaw[1], Shawna Zhu[1], Rebecca Chen[1], Abhinaya Venkatesan[1], Roy Lewis[4] and John S. Gilleard[1].

[1] Faculty of Veterinary Science, Host-Parasite Interactions (HPI) program, University of Calgary, Calgary, Canada.
[2]Animal Health Diagnostic Center, Cornell University, USA.
[3]Department of Veterinary Sciences, University of Wyoming, United States.
[4]Merck Animal Health, Calgary, Alberta, Canada

Presenting author: venkateswararao.pari@ucalgary.ca

KEYWORDS
Bovine Eimeria, metabarcoding, COI mtDNA, Assign Taxonomy, IDTAXA, BLAST-97

ABSTRACT

Bovine coccidiosis is an important diarrheal disease of cattle caused by thirteen *Eimeria* species of varying pathogenicity. Conventional diagnosis of oocysts by morphology has problems with overlapping morphology for a few closely related species. Therefore, we are developing metabarcoding approaches to investigate bovine *Eimeria* diversity from fecal oocyst DNA targeting 18s rDNA and COI mtDNA markers. Morphologically validated ethanol fixed bovine *Eimeria* oocysts from >100 clinical samples were collected from dairy cattle and morphological species identification performed on ~100 oocysts per sample. Partial COI mtDNA (459 bp) was PCR amplified, sequenced at depth using the short-read Miseq (Illumina) platform. The raw sequence data was quality filtered using the DADA2 bioinformatic pipeline to generate 708 Amplicon sequence variants (ASVs). The ASVs were further filtered to 97 with the parameters of sequence length of 350 -500 bp and a minimum read depth of 200. The filtered ASVs were further species classified with three classification pipelines namely 1. The Bayesian classifier- Assign taxonomy 2. IDTAXA-60 with 60 percent threshold, IDTAXA-30 with 30 percent threshold and 3. BLAST-97 (> 97 percent identity). The BLAST classification was modified to assign sequences to species level if $\geq$ 97 percent identity to references otherwise they were assigned to genus. The BLAST-97 percent threshold was chosen based on an analysis of within and between species variation of available bovine *Eimeria* COI reference sequences. A bespoke apicomplexan COI species database was constructed using Geneious with 2664 sequences from MIDORI and GenBank databases containing 24 sequences from 10 bovine *Eimeria* species for use in the Assign taxonomy and IDTAXA pipelines. Both BLAST-97 and Assign Taxonomy classified 61 (70.9 %) ASVs to species level, IDTAXA-30 classified 52 (60.5%) and IDTAXA-60 classified 16 (18.6 %) ASVs to species level. The BLAST-97 pipeline classified most bovine *Eimeria* species (8) along with *Eimeria tenella* (positive control) followed by assign taxonomy classifying 8 bovine *Eimeria* and not *Eimeria tenella.* The IDTAXA-30 pipeline classified 6 bovine *Eimeria* and did not classify *Eimeria zuernii* and *Eimeria tenella.* IDTAXA-60 pipeline classified *E. alabamensis, E. bukidnonensis, E. brasiliensis, E. cylindrica* and *E. illinoisensis. E. bovis, E. zuernii, E. auburnensis* and *E. cylindrica* molecular species assignments by BLAST-97 and Assign Taxonomy correlated well (>0.5 R^2) with morphological species identification compared to three bovine *Eimeria* for IDTAXA-30. None of the four bovine *Eimeria* IDTAXA-60 species assignments correlated with morphology species identification.

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[123], Cody Turner[123], Luke Saville[123], Yubo Cheng[123], Travis Haight[123], Riya Roy[1234], Gerlinde A.S. Metz[24], Majid Mohajerani[4], Athanasios Zovoilis[1234]
[1] Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, Alberta, Canada
[2] Southern Alberta Genome Sciences Centre, Lethbridge, Alberta, Canada
[3] Canadian Centre for Behavioral Neuroscience, Department of Neuroscience, Lethbridge, Alberta, Canada
[4] Alberta RNA Research and Training Institute, Lethbridge, Alberta, Canada

Presenting author: liam.mitchell@uleth.ca

KEYWORDS
RNA modification, epi-transcriptomics, inosine, repetitive elements, bioinformatics

ABSTRACT
Alzheimer's disease (AD) is a multifactorial severe neurodegenerative disorder, whose underlying molecular pathology is largely unclear. Editing of RNA involving conversion from adenosine to inosine (A-to-I) has been connected with cellular function, bringing epi-transcriptomics to the spot-light. 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, and we report that changes in A-to-I editing of SINE B2 RNAs is associated with epi-transcriptome response to amyloid beta neuro toxicity and pathology. We show that A-to-I editing at specific B2 RNA positions is increased 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 regulation of gene expression 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. To this end, inhibiting A-to-I editing activity in hippocampal cells results in increased destabilization of B2 RNAs. 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).

Gene by environment interaction study of major depressive disorder and peer victimization in a pediatric population

Min Jae Kim[1], Tracy Vaillancourt[2], Heather Brittain[2], Paul Arnold[1]
[1] University of Calgary, Calgary, Canada
[2] University of Ottawa, Ottawa, Canada

Presenting author: minjae.kim@ucalgary.ca

KEYWORDS
Genetics, Mental Health, Depression, Environmental Influences on Health, Pediatrics

ABSTRACT
Childhood adversity such as peer victimization (i.e., bullying) is a common phenomenon, where the experience can consist of physical, verbal, indirect and cyber-related violence. Children who have experienced peer victimization have an increased susceptibility to various psychiatric disorders including major depressive disorder, and therefore it has been identified to be a risk factor for depression. However, the effect of peer victimization varies between individuals, and therefore further research is required to understand how genetic predisposition in conjunction with other environmental factors interacts with peer victimization to confer risk for depression. The objective of the study is to identify whether polygenic risk will be associated with the development of depression in conjunction with childhood peer victimization. Data from a longitudinal McMaster Teen Study have been obtained, where students were initially assessed in Grade 5 (n=875) and have been followed to age 22. Peer victimization data have been measured using the Indirect Aggression Scale Target Version and depression symptoms were measured using the Behavioral Assessment System for Children-2. We will use genome-wide single nucleotide polymorphism (SNP) data to identify the genetic variants associated with depression as a quantitative trait. History of childhood peer victimization will be included as a covariate to investigate gene-by environment interactions associated with depressive symptoms. Lastly, polygenic risk scores for depression from previously published data will be obtained to test whether they predict susceptibility to depression in children faced with childhood peer victimization.

Identifying the impact of editing on B2 RNA stability: Connecting the dots between editing and cellular response to stress

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

(1) Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, AB, Canada
(2) Southern Alberta Genome Sciences Centre, University of Lethbridge, Lethbridge, AB, Canada
(3) Canadian Centre for Behavioral Neuroscience, University of Lethbridge, Lethbridge, AB, Canada

Presenting author: riya.roy@uleth.ca

KEYWORDS

B2 SINE RNAs, Adenosine-to-Inosine editing, Heat shock, Transcription, Mouse fibroblasts

ABSTRACT

Short Interspersed Nuclear Elements (SINEs) are one of the most abundant classes of non-coding RNAs involving murine B2 and human Alu as the most frequent ones. Adenosine-to-Inosine RNA editing by adenosine deaminase acting on RNA (ADAR) enzymes, is a very common post-transcriptional modification that predominantly targets double stranded RNA that majorly includes B2 and Alu SINEs and deaminates adenosines to form inosines via a hydrolytic deamination reaction in those specific duplex regions. After this posttranscriptional process, inosine is recognized as guanine by the cellular machinery. Recent studies suggest that the mouse B2 RNAs are capable of binding and repressing RNA polymerase II during pre-stress condition and prevents the mRNA transcription of stress related genes. Upon the detection of stress however, the B2 RNA is cleaved, allowing for stress related genes to be transcribed [1,2]. In spite of this finding, the role of Adenosine-to-Inosine (A-to-I) editing in B2 SINE RNA stability remains unclear. Here, we are investigating the connection between SINE RNA editing and their stability. In particular, the correlation between SINE RNA processing ratio and A-to-I editing during response to cellular stress in mouse fibroblasts.

REFERENCES

1. https://doi.org/10.1016/j.cell.2016.11.041
2. https://doi.org/10.7554/eLife.61265

A transcriptomics approach to understand the interactions between Escherichia coli and copper, silver and gallium metal ions

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

[1] University of Calgary, Calgary AB, Canada.

Presenting author: danielarmando.salaza@ucalgary.ca

KEYWORDS
Transcriptomics, copper, gallium, silver, *Escherichia coli*

ABSTRACT
In the current antimicrobial resistance era, where resistance to most antibiotics now common, metal and metalloid based antimicrobials (MBA) have re-surfaced as an alternative to handle infectious agents. A variety of MBAs are under research for their antimicrobial activities, with silver and copper products already in the market. However, the precise mechanisms of action behind them are not fully established. Here we explore the cell response profile of *Escherichia coli* when challenged to grow under the presence of silver nitrate, copper sulfate and gallium nitrate. Growth curves using a gradient of metal salt concentrations were run to determine a subinhibitory concentration for each of Ag+1 (10 µM), Cu+2 (39 µM) and Ga+3 (1250 µM). The chronic challenge took place in M9 minimal media + glucose and 1 x 10^9 cells were pelleted after 10 hours to perform total RNA extraction. After rRNA depletion and cDNA libraries creation, RNAseq was performed by means of Illumina MiSeq 150 (2.5M read pairs per sample, 3.8Gbp) with an overall >Q30 of 94.85%. Bioinformatic processing of data is underway using the rsubread and DESeq2 Bioconductor packages in R-Studio, and preliminary findings from the data analysis are expected to be presented. This will complement previous findings from a series of toxicogenomic screenings performed by our group years before [1-3].

REFERENCES

1. https://doi.org/10.1093/mtomcs/mfab071
2. https://doi.org/10.3390/genes10010034
3. https://dx.doi.org/10.3390%2Fgenes9070344

Dissecting the brain tissue specific expression of circular RNAs through Nanopore Sequencing

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

1 University of Lethbridge, Lethbridge, Canada

Presenting author: haight@uleth.ca

KEYWORDS
circRNA, Nanopore, Transcriptome, Non-Coding RNA, Rolling Circle Amplification

ABSTRACT
Circular RNAs are a recent addition to the class of non-coding RNAs and are generally characterized by their lack of a 5' cap and a 3' polyadenylated tail. The expanding research into circular RNA biogenesis and function implies that they are simply not splicing errors as previously thought but hold conserved biological functions contributing to regulation of gene expression. However, we currently lack a comprehensive picture of the circular transcriptome of the different regions of human and mouse brain. Here we utilize a method based on Nanopore Sequencing for the detection and analysis of circular RNA in order to determine how circRNA expression differs across mouse and human brain tissues.

Parasites and polyjuice: the role of molecular mimicry in modulating the host immune system

Kaylee D. Rich[1], Viraj Muthye[1], James D. Wasmuth[1]

[1] University of Calgary, Calgary AB, Canada

Presenting author: kaylee.rich@ucalgary.ca

KEYWORDS
Molecular mimicry, sequence analysis, immune evasion, convergent evolution, high-throughput bioinformatics, global health pathogens

ABSTRACT
Successful parasitic organisms must survive the onslaught of their host's immune system. Modulation of the immune system to ignore the parasite is a common mechanism though only few examples are known of parasite molecules involved in this manipulation [1-3]. Here, we set out to identify new potential immune regulators by focusing on convergent evolution, in which parasite proteins have evolved to mimic the structure and/or sequence of host proteins. First, we have taken a fragmented-sequence approach [4], building a pipeline that identifies short peptides that are unexpectedly shared between the parasite and the host, i.e. they are missing in more closely-related species. We searched the proteomes of parasitic eukaryotes, and pathogenic bacteria and fungi. We found relatively few potential mimics in most pathogenic bacteria, with most species only having 5-to-20 hits. An exception was *Mycobacteria*, with 100+ potential mimics. We found more hits in the pathogenic fungi and parasitic eukaryotes. Of note, when we compare the intracellular *Trypanosoma cruzi* [5] and the extracellular *Trypanosoma brucei* [6], we found considerably more proteins containing potential mimics in the former: 1230 vs 191. Currently, we are determining the functions of these potential mimics and divining their likely role in host-parasite interactions. We plan to extend our searches to structural similarity, which we expect will increase the sensitivity of detection.

REFERENCES

1. https://doi.org/10.1038/nm1734
2. https://doi.org/10.1073/pnas.160027697
3. https://doi.org/10.1074/jbc.M502511200
4. https://doi.org/10.1371/journal.pone.0017546
5. https://doi.org/10.1038/s41598-018-32877-2
6. https://doi.org/10.1126/science.1112642

Performance assessment of long-read assemblers using nematode species with different genome natures

Grace M. Mariene[1], Dr James D. Wasmuth[1]

[1] University of Calgary, Calgary, Canada

Presenting author: mcessie2010@gmail.com

KEYWORDS
Genomics, long reads, heterozygosity, repeats, contiguity

ABSTRACT
Genomics work evolves rapidly. Fragmentation of short-read assemblies has compelled the need for longer and more accurate reads leading to better quality genome assemblies. As a result, more assembly programs are developed and benchmarked using various organisms. However, most benchmarking does not consider organisms of the same phylum with contrasting genome natures. As many emerging projects now target the assembly of non-model organisms with different genome complexities, it is important to understand how different assemblers work and perform on these different genomes. We have showed how different genome natures can affect the assembly quality when using different assembly programs. To achieve this, we followed the published *Caenorhabditis bovis* protocol to assess the performance of four commonly used long-read assemblers using two study systems whose data is available for comparison: the smaller and less complex *Caenorhabditis bovis* generated using the Oxford Nanopore MinION and the larger and more complex *Haemonchus contortus* generated using the PacBio RSII platform. The *C. bovis* has a smaller genome size, lower number of repeats and level of heterozygosity as compared to the *H. contortus*. First, we generated preliminaries assemblies using the different assembly programs then carried post-processing polishing steps using various long- and short-read polishing programs. The quality of the draft assemblies generated was tested for contiguity and accuracy using different commonly used metrics; the assembly size, number of fragments, N50, size of the longest contig and the BUSCO score. We then carried out genome-to-genome alignments to ascertain the genome differences between the generated draft assembly and its reference genome. In addition, we ran the Inspector program to map the long reads on the draft assemblies and quantify structural and small-scale errors for each generated assembly. For the less complex organisms, De-Bruijn Graph assemblers, Redbean and Flye performed better than the Overlap-Layout-Consensus (OLC) assemblers, SMARTdenovo and Canu. For complex organisms, however, Redbean which uses the Fuzzy DBG algorithm and the Canu and SMARTdenovo OLC assemblers yielded better assemblies than Flye which uses the generalized DBG algorithm. Overall, the Redbean assembler dealt with repeats best for both genome natures. In general, there is no single gold standard assembler that outperforms all others across species. For this reason, it is recommended to test a few assemblers; two or more, before settling on one. Our study guides the choice of an assembler given the nature of the genome of species of interest and overall research goal.

Species-specific prediction of glycogen-targeting enzyme activities in dominant vaginal *Lactobacillus* species

Kevin Muirhead[1], Karen V. Lithgow[2], Shaelen Konschuh[2], Alison C. Roxby[3-6] and Laura K. Sycuro[1,2,7-9]

[1]Department of Biochemistry and Molecular Biology, University of Calgary, Calgary CANADA
[2]Department of Microbiology, Immunology and Infectious Diseases, University of Calgary, Calgary CANADA
[3]Department of Global Health, University of Washington, Seattle USA
[4]Department of Medicine, University of Washington, Seattle USA
[5]Department of Epidemiology, University of Washington, Seattle WA USA
[6]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle WA USA
[7]Calvin, Phoebe & Joan Snyder Institute for Chronic Diseases, Cumming School of Medicine, University of Calgary, Calgary CANADA
[8]Alberta Children's Hospital Research Institute, Cumming School of Medicine, University of Calgary, Calgary CANADA
[9]International Microbiome Centre, University of Calgary, Calgary CANADA

Presenting author: Kevin Muirhead kevin.muirhead@ucalgary.ca

KEYWORDS

microbiome, women's health, carbohydrate active enzymes, metagenomics, microbiology

ABSTRACT

An optimal vaginal microbiome is characterized by one or two *Lactobacillus* species that dominate the vaginal niche and produce lactic acid to prevent colonization by pathogens. *Lactobacillus* dominance is thought to be supported by glycogen, an abundant host-derived carbohydrate known to contribute to the fermentative production of lactic acid. The type I pullulanase (*pulA*) gene encodes a key bacterial enzyme targeting glycogen's branched alpha 1,6 glycosidic linkages, releasing catabolites for microbial consumption. *L. crispatus* produces a secreted pullulanase enzyme that has been shown to be critical for this species' growth on glycogen. We hypothesize that the presence of a functional pullulanase gene may be important for sustained *Lactobacillus* dominance, but it remains unknown whether other dominant vaginal *Lactobacillus* species encode secreted pullulanases or other glycogen-targeting enzymes. We created a species-specific carbohydrate active enzyme (CAZyme) database by applying homology, HMM and kmer based prediction tools (dbCAN and eCAMI) to vaginal *Lactobacillus* pangenomes. The GH13 family, containing most glycogen-targeting enzymes made up 10-40% of the glycoside hydrolase enzymes. *L. crispatus*, widely considered optimal for health, had the lowest fraction of GH13 enzymes (21/192, 10.94%) and the greatest diversity of other CAZyme families. In contrast, *L. iners*, a prevalent but suboptimal species, had the highest fraction of GH13 enzymes (20/49, 40.82%). Of six *Lactobacillus* species screened, only *L. crispatus* and *L. iners* were predicted to encode secreted pullulanase enzymes. Upon examining the presence of the *pulA* gene within 115 *L. crispatus* isolate genomes and metagenomes, we showed that 30% of strains lack the *pulA* gene or encode a functionally inactivated allele. To start to understand the impact of *pulA* inactivation on community stability and health *in vivo*, we conducted a pilot study of African adolescents (N=17) in which *L. crispatus*-dominated samples lacking *pulA* exhibited 66% lower pullulanase activity and 40% lower D-lactic acid levels than those with functional *pulA*. Altogether, this work suggests that directly accessing glycogen via a secreted

pullulanase enzyme may contribute to the high prevalence of *L. crispatus* and *L. iners* by providing these species a boost in protective lactic acid production. However, our work also suggests producing the *pulA* gene could have a fitness cost, driving some strains to 'cheat' off co-colonizing *pulA* competent strains. Learning how to therapeutically exploit *L. crispatus' pulA* gene, as well as the other carbohydrate-catabolising pathways specific to this species, could lead to the development of next-gen probiotics and prebiotics that improve women's health.

Stabilized Marker Gene Selection for single-cell RNA-seq data

Sandesh Acharya[1], Pathum Kossinna[2], Jiami Guo[3], Qingrun Zhang[4]

[1]:Department of Medical Science; [2]: Department of Biochemistry and Molecular Biology; [3]: Department Of Cell Biology and Anatomy; [4]: Department of Mathematics and Statistics; University of Calgary, Alberta, Canada.

Presenting author: sandesh.acharya@ucalgary.ca

KEYWORDS
Marker Genes, LASSO Regression, Co-expression analysis, Astrocytes, SCOPE

ABSTRACT
Background:

Most computational methods in scRNA-Seq analysis can discover cell types that differ in transcriptome and function (e.g. a sophisticated clustering tool). For different cell types, the methods usually identify a marker (gene) that is uniquely expressed in the group as its marker, leaving the in-depth functional analysis as a relatively independent future work. Certainly, single markers would not be an ideal starting point to understand the functions; alternatively, too many markers identified using naive regression models will lead to a daunting number of markers that are difficult to understand or experimentally tested. As such, we need a method that can produce a reasonably small number of markers that are feasible for experimental validation.

In addition, currrent state-of-the-art methods for marker gene identification involve single-gene-based DE analysis. However, it is natural to expect that complex models analyzing multiple genes jointly should lead to in-depth understanding of cellular functions. However, methods employing such complex models are unstable and generate inconsistent results. Thus, we need a method that not only identifies small number of marker genes, but also generates consistent results across different datasets.

Methods:

In this work, we propose the implementation of SCOPE (Stabilized Core gene Identification and Pathway Election) in single-cell RNA-seq data. The method is build on top of various Clustering tools available for sc-RNAseq data. The input to the method is normalized expression data along with cluster labels (generated by any clustering algorithm). Then, treating each cluster as a separate cell type, LASSO (Least Absolute Shrinkage and Selection Operator) is employed to identify core genes that are underlying the unique functions of each cell types. Correlation and enrichment analyses will be followed to identify surrounding genes. The results from DE (Differential Expression), Co-expression and Pathway Analysis are taken together to identify the marker genes. Results:

Two levels of analysis are done. First, SCOPE has been implemented on various gold-standard sc-datasets (GSE109999, GSE81861, Mouse PBMCs and GSE75748) to identify the marker genes for different cell types in each dataset. Next, SCOPE is implemented on in-house generated cortex scRNA-seq data to identify different types of astrocytes and their marker genes in mouse brain. Our results show that SCOPE-identified marker genes and pathways are highly accurate as well as consistent.

Significance:

This framework incorporates optimizations brought by multiple regression and gene-gene interactions, whilst retaining stability in large part due to two levels of stabilizations (Iterative LASSO Regression and Co-expression analysis).

REFERENCES

1. https://www.biorxiv.org/content/10.1101/2021.12.21.473727v1.full
2. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6105007/
3. https://pubmed.ncbi.nlm.nih.gov/28319088/
4. https://pubmed.ncbi.nlm.nih.gov/27534536/

Input File Read Order Affects the Reproducibility of Structural Variation Calling from Long-Reads

Kyle Lesack, James Wasmuth
University of Calgary, Calgary, Canada

Presenting author: kyle.lesack1@ucalgary.ca

KEYWORDS

structural variation, reproducibility, benchmarking, bioinformatics, genomics

ABSTRACT

INTRODUCTION
A major source of genetic diversity is large structural variations (SVs) in an organism's chromosomes. SVs have considerable impacts in a variety of evolutionary processes, including adaptation and speciation. We and others compared SV identification methods and have shown significant variation in their accuracy [1][2]. For short-read data, how software handles ambiguous read-to-genome mappings is a surprising and significant source of variation in SV identification; changing the order of the reads in the mapping file led to changes in predicted SVs [3]. It is widely thought that the identification of SVs would improve with the use of long-read data, from PacBio and Oxford Nanopore. However, we were surprised by the variation in methods that used long-read data [citation]. We expanded our analysis to identify the impact of read order using three popular long-read callers.

METHODS
PacBio sequencing data were obtained for 15 *Caenorhabditis elegans* isolates from the *C. elegans Natural Diversity Resource* database [4]. For each sequencing run, we created five additional FASTQ files with randomized orders of the same reads. We predicted structural variants for each FASTQ file using three tools: pbsv [5], Sniffles [6], and SVIM [7]. For each isolate, we compared the predictions generated from the different FASTQ files and identified unique variant predictions and SVs with discordant breakpoints.

PRELIMINARY RESULTS
The FASTQ file read order affected the total number of predicted variants and breakpoint identification in each caller. Sniffles was more resilient to FASTQ shuffling (median discordant SVs called from the original FASTQ and shuffled FASTQs = 16; median discrepant breakpoints called from the original FASTQ and shuffled FASTQs = 55.5). SVIM was the most susceptible to FASTQ shuffling (median discordant SVs called from the original FASTQ and shuffled FASTQs = 34.5; median discrepant breakpoints called from the original FASTQ and shuffled FASTQs = 12267.5)

CONCLUSIONS
For all callers, the FASTQ read order affected both the quantity of SV predictions and their breakpoints. Understanding the causes of these discrepancies will allow tool developers to develop deterministic methods that will lead to improved reproducibility of future SV studies.

REFERENCES

1. https://doi.org/10.1101/2022.03.11.483485
2. https://doi.org/10.1186/s13059-019-1720-5
3. https://doi.org/10.1093/bioinformatics/btw139
4. https://doi.org/10.1093/nar/gkw893
5. https://github.com/PacificBiosciences/pbsv
6. https://doi.org/10.1038%2Fs41592-018-0001-7
7. https://doi.org/10.1093/bioinformatics/btz041

Identifying somatic mutations from DNA derived from stereo-EEG electrodes in patients with focal cortical dysplasia

Mascarenhas R[1], Khanbabaei M[1], Ghaderi N[1], Serrano C[1], Fatehi M[1], Rosin JM[1], Calgary Comprehensive Epilepsy Program Collaborators, Liu Y[1], Hader W[1], Tarailo-Graovac M[1], Klein KM[1].

[1] University of Calgary, Calgary, Canada.

Presenting author: rumika.mascarenhas@ucalgary.ca

KEYWORDS
Epilepsy, Focal cortical dysplasia, Stereo-EEG electrodes, Somatic mutations, Variant calling.

ABSTRACT

Background
Focal cortical dysplasia (FCD) is one of the common causes of epilepsy and is characterized by abnormal brain cell organization and development. Studies have investigated brain tissue obtained from patients undergoing epilepsy surgery and have recognized somatic mutations as a cause of FCD. However, with this classical method of identifying brain somatic mutation, access to healthy brain tissue for comparison is limited. To overcome this limitation, we aim to detect somatic mutations by exploiting trace brain DNA obtained from stereo-EEG (SEEG) electrodes as described in a recent study [1]. The broad coverage of SEEG implantation across multiple brain lobes allows us to compare somatic variants in abnormal and healthy brain regions. We have further optimized this method by sorting nuclei into neurons, astrocyte, and microglia to eliminate contaminating cells.

Methods
SEEG electrodes were collected in tubes containing phosphate-buffered saline according to groupings based on brain regions. Cells were extracted from the electrodes via centrifugation and subjected to nuclei isolation (Invent Biotechnologies). Isolated nuclei were stained with DAPI (DNA stain) and fluorescent-labelled with Anti-NeuN (neurons), Anti-LHX2 (astrocytes) and Anti-SPI1 (microglia) for fluorescence-activated nuclei sorting (FANS). Following sorting of nuclei, DNA amplification was performed using primary template amplification (Bioskryb Genomics). Amplified DNA was purified and then quantified using Qubit™ fluorometer. Short tandem repeat (STR) multiplex assay was performed on amplified DNA samples and blood from the patient.

Results
In the samples processed so far, FANS resulted in an average of 406 neuronal nuclei, 201 astrocyte nuclei and 87 microglia nuclei per region from electrodes across 19 brain regions obtained from 5 patients. Following amplification and purification of selected nuclei, DNA quantification yielded up to 6.7ug. STR analysis confirmed the presence of identical human DNA in blood and amplified DNA samples.

Conclusion
Using nuclei isolation and FANS we have been able to successfully sort nuclei into neurons, astrocyte, and microglia. DNA yield from amplified DNA samples shows that sufficient DNA concentrations can be obtained from sorted nuclei for downstream genetic analysis. We intend on performing deep exome sequencing on DNA samples obtained from SEEG electrodes and carry out bioinformatic analyses to

detect and prioritize somatic variants in patients with FCD. We hope that using our unique approach we can determine novel FCD-causing genes that can help uncover the underlying mechanism and can pave a path for development of better treatments for patients with FCD in the future.

REFERENCES

1. https://pubmed.ncbi.nlm.nih.gov/31489630/

Micropropagation and Transformation of *Cannabis sativa* L.

Gregory Robinson[1], Igor Kovalchuk[1]
[1] University of Lethbridge, Lethbridge, AB, T1K3M4, Canada

Presenting author: g.robinson@uleth.ca

KEYWORDS
Micropropagation, Plant Biotechnology, Transformation, *Cannabis sativa,* Regeneration

ABSTRACT
Cannabis sativa L. is a medicinal plant that has been used for thousands of years, however, due to decades of prohibition and stigmatization, little research has been performed on propagation and transformation techniques. Micropropagation is an alternative crop reproduction method where plants can be aseptically propagated for *Cannabis* plant multiplication and enables gene editing to be performed. Various tissue culture protocols for *Cannabis* have been reported, however, reports have low efficiency, are controversial, or are limited in tissue types, cultivars and plant growth regulators (PGR) tested. This study examines the regeneration and transformation potential of *Cannabis* leaves, petioles, internodes, nodes, and florets across *Cannabis* cultivars for the use in CRISPR-mediated gene editing. *Cannabis* explants were sterilized using 0.06% sodium hypochlorite & 3% hydrogen peroxide and initiated on Murashige and Skoog agar plates with different concentrations of PGRs to induce callogenesis, shoot induction, and rooting. Transformation of *Cannabis* was performed via EHA105 strain of *Agrobacterium tumefaciens* carrying the pCAMBIA1301 construct with *uidA* gene to test the transformation efficiency by visual inspection of blue staining via GUS assay. Statistical analysis was performed using a chi-squared test or one-way ANOVA followed by Dunnett's post hoc test. Callogenesis, shooting, and rooting of *Cannabis* explants was dependent upon tissue, cultivar, and PGRs. Transformation efficiency was dependent upon explant tissue type. With well-developed micropropagation and transformation techniques, *Cannabis* can be propagated with higher multiplication rates and CRISPR/Cas9 gene editing can be performed to produce new cultivars with optimized or novel traits.

Transcriptomic insights into the parasitic nematode *Heligmosomoides polygyrus*

Stephen M. J. Pollo[1], Aralia Leon Coria[1], Constance A. M. Finney[1], James D. Wasmuth[1]

[1] University of Calgary, Calgary, Canada

Presenting author: stephen.pollo@ucalgary.ca

KEYWORDS
RNA-seq, gene expression, L4, adult, worm

ABSTRACT
Heligmosomoides polygyrus is a parasitic nematode (roundworm) of mice that is closely related to economically important parasites of livestock as well as the hookworm parasites of humans [1]. As a parasite of mice it is more amenable to laboratory manipulations and being maintained in a controlled laboratory environment than its relatives, which must be maintained in their large animal or human hosts. Similar to its close relatives, *H. polygyrus* manipulates its host's immune system during infection to avoid destruction. Among these manipulations is a shift toward the less inflammatory Th2 responses, which has been shown to be beneficial in mouse models of allergy and inflammatory diseases [2]. The worm enters its host during its third larval stage and within the host develops through its fourth larval stage into adults that reside in the lumen of the small intestine to mate and lay their eggs. Unravelling these processes and other processes critical to *H. polygyrus* survival will reveal new targets to use for drug discovery for controlling this group of parasites as well as refine previous predictions of immunomodulatory molecules, which have therapeutic potential. Here, we have used RNA sequencing (RNA-seq) to investigate how the expression of genes varies across the parasite's infection of its host. We targeted the male and female worms separately in four key stages: larval nematodes recently ingested by their host that have encysted in the mouse intestine, larval nematodes that are on the cusp of migrating into the intestine, young adult nematodes which have recently emerged into the intestine, and mature adult nematodes which are feeding in the host intestine and producing eggs. Using differential gene expression analyses we have been able to identify genes important for development and genes important for the males vs the females. We have also identified mechanisms by which male-specific and female-specific gene expression is created and maintained despite the worms having no male-specific or female-specific genes. We uncovered evidence of a switch from aerobic respiration in the larval to stages to anaerobic respiration in the adult stages and hypothesize that aerobic conditions are important for the critical developmental processes of molting and cuticle synthesis. Finally, we have identified genes with potential roles in modulating the host immune system.

REFERENCES

1. https://doi.org/10.1186/s12862-019-1444-x
2. https://doi.org/10.1007/s00281-012-0347-3

Investigating the Efficacy of Cannabis as a Novel Therapeutic for Aggressive Pediatric Brain and Nervous System Tumors

Megan Malach[1], Dongping Li[1], Bo Wang[1], Aru Narendran[2], Igor Kovalchuk[1], Olga Kovalchuk[1]
[1] University of Lethbridge, Lethbridge, Alberta
[2] University of Calgary, Calgary, Alberta

Presenting author: megan.malach@uleth.ca

KEYWORDS
Neuroblastoma, ATRT, Cannabis, Cannabinoids, Endocannabinoid system, pediatric brain tumors, pediatric cancer, pediatric tumors, cisplatin

ABSTRACT
Research into the effects of *Cannabis Sativa* and its anti-cancer effects on a variety of tumors via the endocannabinoid system has grown rapidly since the late 90s [1]. Preclinical data from mostly adult tumor models shows that cannabinoids affect canonical tumorigenic signaling pathways controlling tumor growth, death avoidance, and invasion [2, 3], and that cannabinoid combinations may act more effectively than single cannabinoids via "entourage effects" [4, 5]. Data also indicates that cannabinoids may act in synergy with chemotherapy agents, allowing both a lowered dose and improved symptom control [6, 7]. However, investigation of these effects in pediatric tumors is limited and warrants further investigation [8, 9].

Utilizing existing literature and preliminary data, we hypothesize that the anti-tumor effects of cannabis extracts will be apparent in cell lines derived from NB and AT/RT tumors. We predict that treatment of these cells with whole cannabis extracts will reduce the viability of these cell lines by altering their gene expression to reduce growth and kill the tumor cells while having little or no effect on the viability of normal, non-cancerous cells. We also hypothesize that the effects of these whole extracts can be recapitulated by combining pure forms of the most prevalent cannabinoids into a "reconstituted extract," and that both whole and reconstituted extracts will work synergistically with chemotherapy agents to further reduce tumor growth and affect key tumorigenic signalling pathways. The overarching goal of this project is to investigate the potential and mechanisms of the anti-tumor effects of *C. Sativa* whole extracts in aggressive pediatric brain and nervous system tumors.

Initial results indicate that select extracts slow the growth of NB and ATRT cell lines, but do not affect normal cells. Upon completion of this project we anticipate identifying new cannabis extracts that treat NB and ATRT tumors *in vitro,* act in synergy with chemotherapy agents, and to establish the molecular mechanisms of these extracts as they exert such effects. The increased understanding of how cannabis impacts pediatric cancer viability will assist the future development of new therapies for pediatric patients.

REFERENCES

1. https://pubmed.ncbi.nlm.nih.gov/30116049/
2. https://doi.org/10.1080/15384101.2020.1742952
3. https://dx.doi.org/10.1016/j.tips.2013.03.003
4. https://doi.org/10.1016/j.bcp.2018.06.025
5. https://doi.org/10.1038/npp.2017.209

6. http://dx.doi.org/10.3747/co.23.3487
7. https://pubmed.ncbi.nlm.nih.gov/28560402/
8. https://pubmed.ncbi.nlm.nih.gov/33466435/
9. https://doi.org/10.3390/ph15030359

Proksee: a web server for assembling, annotating, and visualizing bacterial genomes

Jason R. Grant[1], Eric Enns[2], Eric Marinier[2], Arnab Saha-Mandal[3], Chih-Yu Chen[2], Morag Graham[2], Gary Van Domselaar[2], Paul Stothard[1]

[1] University of Alberta, Edmonton, Canada
[2] National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, Canada
[3] University of Manitoba, Winnipeg, Canada

Presenting author: jason.grant@ualberta.ca

KEYWORDS
bacteria, genome maps, visualization, annotation, assembly

ABSTRACT
Proksee (https://proksee.ca) is an analytical toolkit and high-performance genome browser for rapid, reliable and in-depth characterization of bacterial genomes. Input to the program can be in the form of an NCBI accession, GenBank file or collection of FASTA files representing assembled contigs. Support for raw sequence data (FASTQ files) is being added along with a machine learning enabled assembly system. Following sequence input users are presented with a graphical view of the bacterial genome, implemented using a new genome browser called CGView.js (https://js.cgview.ca). This view permits rapid navigation and zooming down to the sequence level. Numerous options are available for customizing the display and for exporting maps and associated information. A growing collection of integrated tools is provided to the user, for performing various analyses that provide insights into gene locations and functions, sequence relationships and other genome properties. The outputs from these tools can be seamlessly added to the graphical view for comparison with other features. Alternatively, results can be examined or downloaded using specialized viewers for tabular data, HTML reports and images.

Proksee is designed to be user friendly, responsive, easy to maintain and extensible. A modern and consistent look and feel is used throughout the system. A responsive design allows the system to function equally well on laptops, tablets and smart phones. Computationally intensive analyses are run as jobs on a multi-server cloud-based system that forms the backend of Proksee. This configuration allows the main interface to remain responsive while a flexible number of worker servers run the integrated tools. A live log shows the progress of each job in real time. Administration of Proksee is simplified by an integrated admin system that provides real-time updates on user numbers, server load and configuration, data submission and running jobs. The popularity of the various tools is also tracked to guide future development.

A framework termed “Prokan” has been created in parallel to the Proksee server, to streamline the addition of new analytical tools. Prokan describes the inputs and outputs for each tool and how the data should be incorporated into Proksee. Over the coming months more tools will be added to Proksee for a variety of analyses including functional annotation, CRISPR identification, taxonomic assignment and metabolic profiling. Coupled with the integrated assembly functionality these will allow Proksee to be a valuable resource to the research community in Canada and abroad.

Large scale genomic analysis of *Escherichia coli* genomes to characterize the transmissible locus of stress tolerance (tLST) and tLST harboring *E. coli*

Peipei Zhang, Xianqin Yang

Agriculture and Agri-Food Canada, Lacombe, Alberta, Canada

Presenting author: Peipei.Zhang@agr.gc.ca

KEYWORDS
tLST, virulence genes, antimicrobial resistance, enterotoxigenic *E. coli*, Shiga toxin-producing *E. coli*

ABSTRACT
A genomic island, the transmissible locus of stress tolerance (tLST) confers *E. coli* the resistance to multiple antimicrobial interventions including elevated temperature, pressure and oxidative stresses[1]. It may contribute to the persistence of pathogenic bacteria in food production and clinical settings and hence poses a threat to public health. It has been in debate whether *E. coli* can harbor both tLST and virulence genes, for example, Shiga toxin-producing genes. Antimicrobial resistance (AMR) of bacteria is another concern to public health. *E. coli* harboring AMR genes and tLST may require greater effort to control. However, information on the co-occurrence of tLST and AMR genes in *E. coli* was lacking.

Therefore, we performed analyses on *E. coli* genomes available (n=18,959) in the database of National Center for Biotechnology Information (NCBI) and investigated the prevalence of tLST and correlations between tLST, and virulence and AMR genes in *E. coli.* The genomes were downloaded and examined for quality using a command line tool, PanACoTA[2]. ClermonTyping[3] was used to identify the phylogenetic group of each genome. Abricate was used to search the relevant genes. The genes collected in ResFinder[4, 5] and VirulenceFinder[5] database were used as references for AMR and virulence genes, respectively.

Four tLST variants were found in 2.7% of *E. coli*, with the most prevalent variant being tLST1 (77.1%) followed by tLST2 (8.3%), tLST3b (8.3%) and tLST3a (6.3%). The majority (93%) of those tLST were in *E. coli* belonging to phylogroup A in which the prevalence was 10.4%. tLST was also found in phylogroup B1 (0.5%) and C (0.5%) but not found in B2 or D-G. The presence of tLST and virulence genes in *E. coli* was overall negatively correlated, but tLST was found in all genomes of a subgroup of enterotoxigenic *E. coli* (genotype ST2322). Of note, no Shiga toxin-producing *E. coli* (n=3,492) harbored tLST. The prevalence of tLST and AMR genes showed different temporal trends over the period 1985 to 2019. However, a substantial fraction of tLST positive *E. coli* harbor AMR genes, posing a threat to public health. In addition, suspicious contaminants in draft *E. coli* genomes deposited in NCBI and false annotated genes in ResFinder database were both found, which could be misleading and should be curated.

In conclusion, this study explored the possibility of using public dataset to answer biological questions and improves our understanding of the genetic characteristic of tLST and *E. coli* harboring tLST.

REFERENCES

1. https://pubmed.ncbi.nlm.nih.gov/34295324/
2. https://doi.org/10.1093/nargab/lqaa106
3. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6113867/
4. https://doi.org/10.1093/jac/dkaa345

5. https://pubmed.ncbi.nlm.nih.gov/24575358/

A Snakemake workflow for detecting Bovine Respiratory Disease pathogens and antimicrobial resistance genes in third generation metagenomic sequencing data

Emily K. Herman[1], Claire N. Freeman[2], Jennifer Abi Younes[2], Dana E. Ramsay[2], Nathan Erikson[2], Paul Stothard[1], Matthew G. Links[34], Simon J. G. Otto[5], Cheryl Waldner[2]

[1] Department of Agricultural, Food, and Nutritional Science, Faculty of Agricultural, Life, and Environmental Sciences, University of Alberta, Edmonton, Canada
[2] Department of Large Animal Clinical Sciences, Western College of Veterinary Medicine, University of Saskatchewan, Saskatoon, Canada
[3] Department of Animal and Poultry Science, University of Saskatchewan, Saskatoon, Canada
[4] Department of Computer Science, University of Saskatchewan, Saskatoon, Canada
[5] HEAT-AMR (Human-Environment-Animal Transdisciplinary AMR) Research Group, School of Public Health, University of Alberta, Edmonton, Canada

Presenting author: eherman@ualberta.ca

KEYWORDS
Bovine, BRD, Genomics, Antimicrobial, Nanopore

ABSTRACT
Bovine respiratory disease (BRD) is a major cause of morbidity and mortality in cattle production. BRD is a complex disease caused by bacterial and viral pathogens, as well as environmental and host factors. While laboratory diagnostics are crucial for informing appropriate antimicrobial use, bacterial culture and antimicrobial susceptibility testing can take up to one week and are thus of limited suitability in BRD detection and treatment. Because rapid treatment is critical to reducing sickness and death from BRD, injectable antimicrobials are often given to cattle at risk of developing disease. Long read metagenomic sequencing has the potential to quickly resolve all nucleic acids in a sample, including pathogen biomarkers and antimicrobial resistance genes (ARGs), and therefore represents a novel strategy for faster diagnostic testing, while also producing a wealth of genomic information from the causative agents of BRD. We have developed a Snakemake bioinformatics workflow for processing metagenomic sequence data produced on Oxford Nanopore Technology's GridION platform from deep nasopharyngeal swabs of cattle. The workflow comprises 22 programs and custom scripts, and performs all aspects of sequence quality control, host sequence filtering, taxonomic classification, ARG detection, identification of genetic elements carrying resistance genes, metagenomic assembly, and serotyping of the BRD pathogen *Mannheimia haemolytica*. In an analysis of feedlot cattle that failed to respond to multiple rounds of BRD treatments, BRD pathogens were detected more frequently by metagenomic sequencing than by bacterial culture. ARGs were detected *in silico*, however with concordance with phenotypic resistance. Metagenomic sequencing has the potential to be a robust, rapid, and evidence-based strategy for cattle producers to guide BRD diagnosis and target antimicrobial treatment.