COMBINE

Australian Computational Biology and Bioinformatics Student Society

ISCB Regional Student Group Australia

Bioinformatics Student Symposium

In-person and virtual conference - 22 Nov 2021

Pawsey

INTERSECT RESEARCH FASTER

ISCB International Society for Computational Biology

SAGC

Melbourne Bioinformatics

COMBINE 2021 Symposium

Welcome

Welcome to the 2021 COMBINE Symposium.

The COMBINE Symposium is a student-run, non-profit conference that aims to gather a diverse group of students and ECRs who are curious, passionate and eager to learn more about bioinformatics and computational biology. This annual event is an opportunity for students and early career researchers to present their work to peers in a relaxed and supportive environment. All attendees are expected to show respect and courtesy to everyone throughout the conference. The ABACBS code of conduct will apply to the symposium.

We hope you enjoy today's Symposium.

Regards, The 2021 COMBINE Symposium Committee.

2021 COMBINE Symposium Committee

Chairs: Aditya Sethi and Sachintha Wijegunasekara

Committee Members: Akari Komori, Alice Whitehead, Giulia Iacono, Himal Shrestha, Ji-Ru Han, Kshitiz Shrestha, Sam Davis, Thomas Litster, Tyrone Chen, Yunwei Zhang, Patricia Sullivan, Pia Campagna

For any enquiries please email symposium@combine.org.au.

COMBINE 2021 Symposium

Symposium Information

Useful links

Symposium zoom link

https://combinesymposium2021.slack.com/

https://www.combine.org.au/symp/combine-symposium-2021/

https://www.abacbs.org/

Follow the #COMBINE21 Symposium on twitter

@combine_au

In-person networking session at Adelaide satellite event

Address: Lecture Theatre G030 on ground floor, Adelaide Health and Medical Sciences Building on North Terrace

Symposium prizes

- Best session talks (first, second and third places)
- Best flash talks (first, second and third places)
- People's choice for session talk
- People's choice for flask talk
- Best tweet using #COMBINE21

eResearch Services

Intersect is a national not-for-profit eResearch support organisation.

Intersect helps researchers increase their impact through expert people, a vibrant member community and innovative technologies. Our services are targeted at enhancing capability (skills) through training and capacity (resources) of organisations, disciplines and research groups through an embedded team of experts.

- Training
- Researcher support
- Research strategy, policy, and procedures
- Researcher engagement
- Programming, scripting, and analytics
- Research application support
- Access to research technology
- Relevant personal expertise... augmented by the Services Team

Membership

Intersect Membership enables universities and research intensive organisations to drive research productivity through enhanced researcher capability, capacity, and more effective integration of local and external digital research infrastructure.

Intersect Membership provides access to:

- Engagement and Support through the placement of an eResearch Analyst
- Training and Education

Additional Services

Any university, research organisation or individual researcher can still access Intersect eResearch services through:

Open Training: Training for Individuals & Teams

Individuals and teams from outside the Intersect memberships can access our most popular training courses delivered online by expert trainers.

eResearch Consulting

Personalised consulting by an eResearch specialist covering various digital technologies including REDCap, NVivo, Excel, SQL, and more.

For more information about Intersect eResearch Services please contact us at eResearch@intersect.org.au

COMBINE 2021 Symposium

Program Overview

The times indicated in the following program are in Sydney, Australia (AEDT) time zone.

10:30 - 10:45 **Symposium Opening**
Chairs: Aditya Sethi and Sachintha Wijegunasekara

10:45 - 12:40 **Genomics Session** INTERSECT RESEARCH FASTER
Chair: Akari Komori
10:50 - 12:20 Session talks
12:20 - 12:40 Flask talks session 1

12:40 - 1:10 *Lunch & Discussion*

1:10 - 1:40 **Keynote Address**
Chair: Aditya Sethi
Transcriptomics: Past, present and future
Prof. Alicia Oshlack, Peter MacCallum Cancer Centre

1:40 - 3:15 **Transcriptomics Session**
Chair: Yunwei Zhang
Session talks

3:00 - 3:15 *Afternoon tea & Discussion*

3:15 - 5:15 **Translational Bioinformatics Session** pawsey
Chair: Fred Jaya
3:20 - 4:50 Session talks
4:50-5:05 Flask talks session 2
5:05-5:15 Discussion

5:15 - 6:15 **Careers Panel Session**
Chair: Alice Whitehead and Akari Komori
Dr Sarah Beecroft, Bioinformatics Applications Specialist, Pawsey supercomputing centre
Etsuko Uno, Biomedical Animator, Walter and Eliza Hall Institute of Medical Research
Dr Thom Quinn, Bioinformatician turned Data Scientist
Dr Sonika Tyagi, Senior Lecturer, Monash University

6:15 - 6:30 **Symposium Closing**
Prizes announcements
Closing address

6:30 **Networking Session**
In-person in Adelaide
Virtual in all other states

COMBINE 2021 Symposium

Invited Speakers

Keynote Speaker

Prof. Alicia Oshlack - (Co)Head of the Computational Biology Program, group leader - Peter MacCallum Cancer Centre

Professor Alicia Oshlack is internationally recognised for her development of bioinformatics methods for a range of applications including single cell RNA-seq, methylation and genomic analysis. Oshlack is involved in many cutting edge collaborative projects utilising high throughput sequencing to investigate disease and development. Oshlack has been recognised by several awards including the Australian Academy of Science, Gani Medal for Human Genetics (2011) and the Georgina Sweet Award for Women in Quantitative Biomedical research (2016).

Careers Panel

Dr Sarah Beecroft - Bioinformatics Applications Specialist - Pawsey supercomputing centre

After PhD completion, Dr. Beecroft obtained a dual fellowship, split between ongoing neuromuscular research and supporting bioinformatics at the Pawsey Supercomputing Research Centre. She now works for Pawsey as a Bioinformatics Applications Specialist. In this role, she works to support bioinformatics users through workflow development and optimisation, training, and advocacy, plus significant collaboration with the Australian BioCommons.

Dr Thom Quinn - Bioinformatician turned Data Scientist

Thom Quinn is a bioinformatician who came to Australia to pursue a PhD in 2016. After completion, he worked as a post-doc with the Applied Artificial Intelligence Institute at Deakin University for a couple of years. Recently, he has taken up a job as a data scientist at a government organization (but still considers himself to be a member of this wonderful and welcoming Australian bioinformatics community). He is looking forward to sharing his experience working as a post-doc and as a data scientist, and to answering questions about the transition away from academia

Dr Sonika Tyagi - Senior Lecturer - Monash University

Dr. Sonika Tyagi has a Ph.D. in Computational Biology and over 15 years of work experience in academia & industry. In 2018 she joined a teaching and research position at Monash University to establish her research program. She is currently a Machine Learning lead in the SuperbugAI flagship project. Her expertise is in developing new machine learning tools and pipelines, and apply these methods to solve biological research questions.

Etsuko Uno - Biomedical Animator - Walter and Eliza Hall Institute of Medical Research

Etsuko Uno is a biomedical animator working at WEHI in Melbourne, Australia. Her animations span diverse topics from immunology and gene regulation to cancer and disease. Her works have won numerous awards including the NSF Visualisation Challenge and have been selected by multiple film festivals such as the prestigious SIGGRAPH Computer Animation Festival. She the completed a Masters in Biomedical Science at Rockefeller University, USA. She also holds a Diploma of Animation from RMIT, Melbourne.

COMBINE 2021 Symposium

Detailed Program

Symposium Opening - 10:30am - 10:45pm

AEDT — *Chairs: Aditya Sethi and Sachintha Wijegunasekara*

10:30 - 10:45 **Acknowledgement of Country for COMBINE/ABACBS/Phylomania**
Denis

Welcome to COMBINE 2021
Aditya Sethi and Sachintha Wijegunasekara

Genomics Session - 10:45am - 12:40pm

Chair: Akari Komori

10:50 - 11:05 **Genomic landscape of diversification, selective sweeps, and demographic history of an anthroponotic parasite**
Swapnil Tichkule

11:05 - 11:20 **An automated meta-caller to detect de novo mutations from whole genome trio data using cloud computing technology**
Anushi Shah

11:20 - 11:35 **PhiloBacteria: A new tool to infer phylogenetic trees from recombinant bacterial genomes**
Nehleh-Fatemeh kargarfard

11:35 - 11:50 **Developing a novel CRISPR-Cas9 based method for characterisation of bacterial pathogens using Oxford Nanopore sequencing**
Hugh Cottingham

11:50 - 12:05 **AliSim: Phylogenetic Sequence Simulator in the Genomic Era**
Ly Trong Nhan

12:05 - 12:20 **Is the Human genome in mutation equilibrium?**
Katherine Caley

12:20 - 12:40 **Flask talks session 1**
Chelsea Matthews, Amarinder S. Thind, Holly A. Withers, Urwah Nawaz

Lunch & Discussion 12:40pm - 1:10pm

Keynote address - 1:10pm - 1:40pm

1:10 - 1:40 *Chair: Aditya Sethi*

Transcriptomics: Past, present and future
Prof. Alicia Oshlack, Peter MacCallum Cancer Centre

Transcriptomics Session - 1:40pm - 3:15pm

Chair: Yunwei Zhang

1:45-2:00 **RNA-seq regulatory network inference revealed an association between transcription factor SETX and neurogenerative pathways under prolonged autophagy induction**
Wenjun Liu

2:00-2:15 **A universal naming system for human alternative splicing**
Angelita Liang

2:15-2:30 **Identification, classification, and prioritization of most influential players in normal biological processes and diseases**
Adrian (Abbas) Salavaty

2:30-2:45 **scFeatures: automatic feature generation for single-cell and spatial data**
Yue Cao

2:45-3:00 **Using multimodal single cell data to predict regulatory gene relationships and to build a computational direct cell reprogramming model**
Andy Tran

Afternoon tea & Discussion 3pm - 3:15pm

Translational Bioinformatics Session - 3:15pm - 5:15pm

Chair: Fred Jaya

3:20-3:35 **Integrating genomic location and sequence contexts of point mutations to improve classifying cancer tissue of origin**
Anh Phuong Le (Phuong)

3:35-3:50 **CaraVaN: Prioritising Cardiac Variants in the Non-coding genome using boosting algorithm**
Gulrez Chahal

3:50-4:05 **Genomic epidemiology of Clostridioides difficile in Australia, 2013-2018**
Keeley O'Grady

4:05-4:20 **Landscape genetics to determine factors for ongoing transmission of onchocerciasis in the transition region of Ghana**
Himal Shrestha

4:20-4:35 **disTIL: a turnkey approach to profile the immune landscape in cancer**
Rachel Bowen-James

4:35-4:50 **Investigating how transcriptional plasticity drives drug resistance in cancer by combining computational modelling with single cell genomics**
Alejandro Casar

4:50-5:05 **Flask talks session 2**
Mikhail Dias, Boris Ka Leong Wong, Sanghyun Lee, Ann Rann Wong, Holly Martin

5:05-5:15 **Discussion**

Careers Panel Session - 5:15pm - 6:15pm

Chairs: Alice Whitehead and Akari Komori

5:15-6:15

Dr Sarah Beecroft, *Bioinformatics Applications Specialist*, Pawsey supercomputing centre
Etsuko Uno, *Biomedical Animator*, Walter and Eliza Hall Institute of Medical Research
Dr Thom Quinn, *Bioinformatician turned Data Scientist*
Dr Sonika Tyagi, *Senior Lecturer*, Monash University

Symposium Closing - 6:15pm - 6:30pm

6:15 - 6:30

Chairs: Aditya Sethi and Sachintha Wijegunasekara

Prize presentations
Closing remarks

Networking Session 6:30pm onwards

COMBINE 2021 Symposium

Sponsors

Symposium Principal Sponsor

Pawsey Supercomputing Research Centre

The Pawsey Supercomputing Research Centre, an unincorporated joint venture between CSIRO – Australia's national science agency, and WA public universities, is accelerating scientific breakthroughs through world-class expertise and infrastructure in supercomputing, cloud services, data-intensive analysis, storage and visualisation. From speeding the development of new crops to feed us into the future to developing improved diagnostic tests for coronaviruses, Pawsey's supercomputers are already being used to help accelerate some of Australia and the world's most important research projects across medical research, artificial intelligence, chemistry, physics, engineering, and more. The Centre is supported by the Australian and Western Australia governments and its partner organisations.

Contact: help@pawsey.org.au

https://pawsey.org.au/

Symposium Session Sponsor

Intersect Australia

Intersect Australia is a leading provider of eResearch training, services, and solutions. As a not-for-profit, member-based organisation, Intersect offers training and research support to meet the specific needs of researchers and enhance research productivity. Through a highly experienced team embedded within its member institutions and supported by technical and operational experts, it also offers research-oriented cloud storage, compute, research data management, and research software engineering services.

Contact: eResearch@intersect.org.au

https://intersect.org.au/

General Sponsors

Contact
Ms. Claudia Curcio
E: claudia.curcio@unimelb.edu.au
P: +61 3 9035 5822
https://www.melbournebioinformatics.org.au/

https://www.iscb.org/

https://www.sa-genomics.com.au/index.php

COMBINE 2021 Symposium

Symposium Abstracts

Ten-minute presentations

Listed in alphabetical order of presenters

disTIL: a turnkey approach to profile the immune landscape in cancer

Rachel Bowen-James[1,2], Mark J Cowley[1,3], Chelsea Mayoh[1,3]

1. *Children's Cancer Institute, Lowy Cancer Centre, UNSW Sydney, Kensington, NSW, Australia*
2. *School of Biomedical Engineering, UNSW Sydney, Kensington, New South Wales, Australia*
3. *School of Women's and Children's Health, UNSW Sydney, Kensington, NSW, Australia*

Despite recent advances in immunooncology, our understanding of the factors which determine response to immunotherapy remains extremely limited. This is especially true in childhood cancer: the tumour immune landscape of paediatric cancers remains poorly characterised, meaning the potential for widespread immunotherapeutic treatment of paediatric tumours is uncertain. There is a growing need for large-scale bioinformatic analyses to characterise the immune landscape of paediatric tumours and identify biomarkers of immunotherapy response. We have developed disTIL: a comprehensive bioinformatics toolkit leveraging multi-omic next generation sequencing data to elucidate features of the tumour immune microenvironment and the neoantigen landscape of paediatric tumours. disTIL incorporates algorithms to perform HLA typing, calculation of Tumour Mutational Burden (TMB), analysis of gene expression signatures, neoantigen enumeration and filtering, and cell type deconvolution. disTIL then summarises these results into a single interpretable report to aid clinical decision making. To maximise the portability and flexibility of the toolkit, each module is independently containerised using Docker and implemented in Common Workflow Language (CWL). Here we present disTIL and its application to childhood cancer and precision medicine. Preliminary analyses revealed that despite low overall TMB, all patients thus far have at least five predicted neoantigens presented by their HLA molecules which could be personalised immunotherapy targets. Among these are neoantigens derived from the common and currently undruggable EWSR1-FLI1 fusion found in Ewing's sarcoma. Neoantigens such as these may be shared by several patients, identifying potential targets for off-the-shelf immunotherapy treatments.

Is the Human genome in mutation equilibrium?

Katherine Caley[1], Benjamin Kaehler[2], Von Bing Yap[3], Gavin Huttley[4]

1. *Research School of Biology, ANU College of Science, The Australian National University*
2. *Department of Statistics and Applied Probability, National University of Singapore*
3. *School of Science, University of New South Wales, Canberra, ACT, Australia*
4. *Research School of Biology, ANU College of Science, The Australian National University*

Most models of sequence divergence assume the composition of nucleotides does not change through time. This assumption requires a state of mutation equilibrium which is almost impossible if

the processes affecting mutagenesis change through time. Considerable empirical evidence strongly suggests that this may be incorrect. In my honours, I have addressed this possibility through developing the following statistical measures: a test for the existence of mutation disequilibrium, a test of its equivalence and a measurement of the magnitude of mutation disequilibrium. I used carefully construction of edge cases with simulated data to establish the consistency of the statistics with theoretical expectations. I applied the statistics to empirical data from cases with striking prior evidence for recent perturbations affecting: an entire genome (loss of DNA methylation in Drosophila melanogaster); or, a small genomic segment (FXY in Mus musculus). Using a paired experimental designs, I show the predicted vast excess of small probabilities from the statistical tests. I further show the statistical measure of magnitude is also elevated in these cases. Applying the methods to Human evolution, I conservatively estimate > 50% of our genome is in mutation disequilibrium.

scFeatures: automatic feature generation for single-cell and spatial data

Yue Cao[1,2], Yingxin Lin[1,2], Jean Yee Hwa Yang[1,2], Pengyi Yang[1,2,3]

1. School of Mathematics and Statistics, University of Sydney, Sydney NSW, Australia;
2. Charles Perkins Centre, University of Sydney, Sydney NSW, Australia
3. Computational Systems Biology Group, Children's Medical Research Institute, University of Sydney, Westmead NSW, Australia

Single-cell sequencing and spatial technology enable discovery of cell- and cell type-specific knowledge and has advanced the understanding in many areas in recent years, such as biological systems and diseases. As the data generated by these technologies typically contains tens of thousands of features in the form of gene expression, central to the understanding of such data is the ability to curate and identify useful and interpretable features. Here, we present scFeatures, a tool for automatic feature generation for single-cell and spatial data. With a given data, scFeatures is able to generate 17 classes of features across six major categories of cell type proportions, cell type specific gene expressions, cell type specific pathway scores, cell type specific cell – cell communication scores, bulk gene expressions and spatial metrics. Furthermore, scFeatures implements classification procedure with feature ranking to assess the importance of each feature with respect to the outcome. By applying scFeatures on single-cell RNA-seq, spatial proteomic and spatial transcriptomic data with distinct sample outcomes and cell types, we demonstrate that different classes of features are important to different datasets and discovered insights into diseases.

Investigating how transcriptional plasticity drives drug resistance in cancer by combining computational modelling with single cell genomics

Alejandro Casar[1], David Goode[2], Howard Bondell[3], Luis Lara[4]

1. *The University of Melbourne, School of Mathematics and Statistics, Victoria, Australia*
2. *Peter MacCallum Cancer Centre, Victoria, Australia*
3. *The University of Melbourne, School of Mathematics and Statistics, Victoria, Australia*
4. *Peter MacCallum Cancer Centre, Victoria, Australia*

Transcriptional plasticity is a phenomenon where cells reversibly change molecular states by altering their gene expression. Recent evidence suggests transcriptional plasticity is a key driver of the emergence of drug resistance. We designed a computational model that simulates Acute Myeloid

Leukemia (AML) as a stochastic process at a discrete time resolution. It is based on the rates of DNA mutation, cell division, cell death and switching rates between sensitive and resistant states. We model tumors with initial growth rates between 0.001% and 0.1%; with driver mutation rates between 3.4*10^(-8) and 3.4*10^(-5); testing switching rates between 4.0*10^(-7) and 9.0*10^(-6) from sensitive to resistant, and 0.5 to 50000 times those values for the resistant to sensitive switching rate. Results so far suggest that low rates of transcriptional plasticity have little effect on the growth and development of the cancer pre-treatment. However, as plasticity increases, we can see the tumor develop differently. Significant correlation was observed between tumor fitness, representing the weighted average net growth rate of all the different subpopulations of cells within in the tumor, and the number of resistant cells, which will be explored further. We are following up to understand the interactions between each independent simulation parameter, transcriptional plasticity, and clinical outcomes. To understand how transcriptional plasticity drives drug resistance in cancer, different drug treatment regimens for AML will be simulated and model predictions validated in mouse models of AML using SPLINTR, a synthetic expressed barcoding strategy that allows for cell lineage tracking.

CaraVaN: Prioritising Cardiac Variants in the Non-coding genome using boosting algorithm

Gulrez Chahal[1], Sonika Tyagi[2], Mirana Ramialison[1,3]

1. *Australian Regenerative Medicine Institute, Monash University, Clayton VIC*
2. *Department of Infectious Disease and Monash eResearch Centre, The Alfred Hospital and Monash University, Melbourne, VIC*
3. *Murdoch Children's Research Institute, Parkville VIC*

Congenital heart disease (CHD) are structural and functional defects that occur in the heart during development. There are gene candidates that have been identified to understand CHD. However, in many cases the genetic cause still remains unknown. This can be attributed to the fact that protein-coding genes contribute only ~2% of the genome, while the non-coding genome (~98%) comprises of functional regulatory elements, that are involved in the regulation of the expression of the genes. Recent evidence indicates that variations in these regulatory elements impact gene expression and result in CHD. However, there is no method to investigate non-coding variants in CHD yet. In recent years, several machine learning-based tools have been developed to prioritize these variants in the non-coding genome, however, they are not disease-specific. We present CaraVaN, a cardiac-specific model which annotates and prioritises potential CHD causal variants in the non-coding genome using decision-tree based boosting algorithm. This model learns from cardiac-specific human functional, epigenomic and structural consequence features: histone marks, transcription factor binding sites, 3D chromatin organisation and deleteriousness scores from existing variant assessment tools, to capture non-coding SNVs with potential pathogenicity in CHD. When prioritizing cardiac-pathogenic SNVs, CaraVaN demonstrates an improved performance (ROC AUC=0.704) in comparison to the state of art tools (ROC AUC=0.612). We also validate the performance of CaraVaN to prioritise a functionally known non-coding variant in CHD in chromosome 12. We scored more than 48 million variants in chromosome 12, out of which this variant achieved a high score in the top 8% of these variants. Furthermore, gene ontology (GO) analysis on these top-scoring variants revealed their association with human heart disease phenotypes including atrial

fibrillation and abnormality of cardiovascular system physiology. Overall, CaraVaN is the first tool that evaluates the non-coding variants in CHD and other heart-related diseases.

Developing a novel CRISPR-Cas9 based method for characterisation of bacterial pathogens using Oxford Nanopore sequencing

Hugh Cottingham[1], Louise M. Judd[1], Jane Hawkey[1], Nenad Macesic[1,2], Ryan R. Wick[1], Anton Y. Peleg[1,3], Kathryn E. Holt[1,4]

1. *Department of Infectious Diseases, Central Clinical School, Monash University, Melbourne, Victoria 3004, Australia*
2. *Centre to Impact AMR, Monash University, Melbourne, Victoria, Australia*
3. *Department of Microbiology, Infection and Immunity Program, Monash Biomedicine Discovery Institute, Monash University, Clayton, VIC*
4. *Dept Infection Biology, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK*

Current methods for detection and characterisation of antimicrobial resistant (AMR) bacteria rely on time-consuming, inefficient and costly laboratory-based approaches. Direct-from-sample sequencing approaches would improve turnaround times, however the presence of non-pathogenic DNA frequently overwhelms the signal from the pathogen. CRISPR-Cas9 based enrichment enables selective sequencing of target genomic regions directly from samples. Here, we pilot the use of CRISPR-Cas9 enrichment for detecting and characterising Klebsiella pneumoniae directly from samples using long-read sequencing. We designed guides to target 18 tRNA genes, all seven multi-locus sequence typing genes and three common AMR genes in the Klebsiella pneumoniae genome. We used 21 Klebsiella pneumoniae genomes to test whether our guides would enrich for targeted regions from pure culture. Enriched DNA was sequenced with Oxford Nanopore Technologies and read alignment-based approaches were used to assess if our guides were successful. We found that multiple guides for a single target region were required to provide consistent enrichment - a finding that has not previously been reported. Due to the highly conserved nature of tRNA genes within the Enterobacterales order of bacteria (which are commonly responsible for clinical infections), we found that our guides enriched target regions in six additional genomes from across Enterobacterales. Finally, we spiked in differing abundances of Klebsiella into a mock bacterial community sample and found that our guides successfully enriched target regions in a mixed sample. These results show tremendous promise for CRISPR-Cas9 based enrichment to enable targeted sequencing from clinical samples, potentially allowing for bacterial characterisation with one simple assay.

PhiloBacteria: A new tool to infer phylogenetic trees from recombinant bacterial genomes

Nehleh-Fatemeh kargarfard[1], Aaron Darling[1], Mathieu Fourment[1]

1. *University of Technology Sydney, ithree Institute, Ultimo NSW 2007, Australia*

One of the vital services to maintain public health is detecting and tracking outbreaks of infectious diseases. Phylogenetic tools are one of the most efficient tools to analyze the source, evolutionary history of epidemic outbreaks. According to the rapidly growing number of genomes, it's essential to develop new methods to apply to increasingly large datasets. On the other hand, as asexual organisms, Bacteria have a clonal reproduction; however, their genomes can evolve by different

mechanisms such as point mutation and recombination. When recombination occurs in bacteria genomes, several nucleotides can change together, contributing to considerable evolutionary leaps, for example, helping pathogens develop resistance mechanisms against antibiotics. Practically, it's not only swapping genetic information among organisms, but also it's exchanging their evolutionary histories. Hence, the evolutionary analysis would be different and challenging in the presence of recombination, and ignoring the effect of recombination can result in misleading in phylogenetic. Furthermore, detecting the boundary of recombination events and reconstructing a global tree to illustrate the underlying evolutionary pattern of biological sequences has never been a straightforward problem in terms of accuracy and scalability. We introduce PhiloBacteria, a new tool that uses a hidden Markov model to detect recombination and infer the phylogeny of related organisms simultaneously. Specifically, the algorithm's goal is to reconstruct the true clonal evolutionary history of the organisms while avoiding misleading evolutionary signals originating from recombination events. The method aims at quickly and accurately analyzing large datasets of closely related bacterial genomes. Using simulated data, we investigate the accuracy of our algorithm and compare our results to other established recombination detection methods.

The RNA Atlas: The Splicing "Seq"uel

Angelita Liang[1], Nandan Deshpande[1], Ashwin Unnikrishnan[1], Marc R. Wilkins

1. *UNSW Sydney*

The global understanding of human alternative splicing is incomplete. Currently, the splicing patterns of many cell types as well as the differences in splicing between cell and tissue types remain unexplored. Although the regulatory mechanisms of gene expression and splicing are intimately related both spatially and temporally, their interplay is not fully understood. The RNA Atlas is an ultra-high resolution map of the human transcriptome, encompassing 296 distinct tissues, primary cells and cancer cell lines, generated using ultra-deep short-read polyA capture and total RNA sequencing. We present a comprehensive global view of splicing events and their relative prevalence in non-coding RNAs. We show that tissue and cell types are strongly demarcated by unique splicing signatures, which surprisingly, are restricted to a small number of biological pathways. On the other hand, gene expression signatures are much less well defined, but affect a wide variety of biological pathways. We demonstrate that using splicing patterns with or without gene expression as input into classification algorithms can enhance classification power. Through a value called the "stacking entropy" which measures the information value of biological features, we show that the information content in splicing is fundamentally different to that of gene expression. These findings are recapitulated in the data from the ENCODE project and warrants specialised treatment of quantitative splicing data. Thus, in addition to providing a resource for human alternative splicing, our findings in the Atlas hold many novel insights and broad implications for RNA biology, machine learning as well as clinical diagnostics.

RNA-seq regulatory network inference revealed an association between transcription factor SETX and neurogenerative pathways under prolonged autophagy induction

Wenjun Liu[1], Aaron E. Casey[2,3], Timothy J. Sargeant[4], Stephen M. Pederson[2,3], Ville-Petteri Makinen[1]

1. *Adelaide Medical School, Faculty of Health and Medical Sciences, University of Adelaide*
2. *Computational and Systems Biology, Precision Medicine Theme, South Australian Health and Medical Research Institute*
3. *Australian Centre for Precision Health, Cancer Research Institute, University of South Australia*
4. *Lysosomal Health in Ageing, Lifelong Health, South Australian Health and Medical Research Institute*

Despite the global prevalence of dementia, disease pathogenesis remains unclear, and treatments continue to disappoint. Recent studies found possible associations between neurodegenerative diseases and autophagy dysfunction. To explore this hypothesis, we induced autophagy in three human cell lines using two treatments targeting different stages of the autophagic pathway, collecting RNA-Seq data at three time-points up to 30hrs. Analysis revealed changes in transcriptional activities and biological processes in response to autophagy activation. Although successful activation of autophagy was verified through measuring autophagic flux, the autophagy pathway (KEGG) appeared to be consistently inhibited, rather than activated. This finding suggests that when cells were under prolonged autophagy induction, classical autophagy-related genes may be regulated to maintain autophagic homeostasis. Conversely, the Alzheimer's Disease pathway was consistently activated in response to the extended activation of autophagy. During this work, an existing topology-based pathway enrichment testing method SPIA was improved by incorporating a more robust significance testing strategy. In addition, a novel regulatory network inference method was developed in order to identify the regulatory influences of specific transcription factors on relevant biological pathways. This approach revealed a potential role for the transcription factor SETX on a list of neurodegenerative diseases pathways including AD under prolonged autophagy activation.

AliSim: Phylogenetic Sequence Simulator in the Genomic Era.

Ly Trong Nhan[1], Bui Quang Minh[1]

1. *Australian National University*

Sequence simulations play an important role in phylogenetics. Simulated data is used for many purposes, such as evaluating the performance of phylogenetic methods, hypothesis testing with parametric bootstrap, and, more recently, generating datasets for training machine learning-based applications. However, existing simulators, such as Seq-Gen, Dawg, and Indelible, only support a limited number of substitution models, thus failing to simulate genomic data under complex evolution models. Furthermore, those simulators are computationally intractable for simulating massive datasets with millions of sequences or sites. Here, we introduce AliSim, a new sequence simulator that can efficiently simulate biologically realistic alignments under complex evolutionary models. AliSim supports hundreds of evolutionary models available in the IQ-TREE software, including standard, mixture, and partition models. Additionally, AliSim allows users to simulate alignments that mimic the evolutionary process of a real alignment. AliSim takes only an hour and approximately 1GB RAM to simulate alignments with millions of sequences or sites, while popular software Seq-Gen, Dawg, and Indelible require many hours and tens to hundreds of GB RAM. We

provide AliSim as an extension of the IQ-TREE software version 2.2, freely available at www.iqtree.org.

Genomic epidemiology of Clostridioides difficile in Australia, 2013-2018.

Keeley O'Grady[1], Dr Daniel R. Knight[1,2], Professor Thomas V. Riley[1,2,3,4]

1. *Centre for Biosecurity and One Health, Harry Butler Institute, Murdoch University, Murdoch, Western Australia, Australia*
2. *Marshall Centre for Infectious Diseases Research and Training, School of Biomedical Sciences, The University of Western Australia, Queen Elizabeth II Medical Centre, Nedlands, WA 6009, Australia.*
3. *School of Medical and Health Sciences, Edith Cowan University, Joondalup, WA, Australia.*
4. *Department of Microbiology, PathWest Laboratory Medicine, Queen Elizabeth II Medical Centre, Nedlands, WA, Australia.*

Clostridioides (Clostridium) difficile is an anaerobic, gram-positive bacterial pathogen responsible for a range of intestinal diseases ranging from diarrhoea to life-threatening pseudomembranous colitis. Hardy, highly resistant spores are transmitted by the faecal-oral route before germinating in the intestines. Strains produce one or both of TcdA/TcdB toxins, with some strains also producing an extra toxin – binary toxin or CDT. C. difficile is induced by, and resistant to, a variety of antimicrobials, as gut dysbiosis reduces the commensal microbiota that normally inhibit it. Begun in 2013, the Clostridium difficile Antimicrobial Resistance Surveillance (CDARS) study is Australia's first longitudinal and nationwide surveillance study for Clostridioides difficile infection (CDI). Between 2013-2018, ten diagnostic microbiology laboratories in five states around Australia collected >1500 isolates from human cases of CDI. A representative subset of 492 C. difficile strains underwent Illumina WGS and were analysed in silico for phylogeny, toxins and antimicrobial resistance (AMR). Core-genome Single Nucleotide Polymorphism (cgSNP) analysis with a custom bioinformatics pipeline identified clonal strains that were geographically and temporally dispersed, sometimes by years or thousands of kilometres, suggesting widespread and diverse sources of infection and a role for long-distance transmission of spores. In silico screening for AMR and toxin genes revealed a susceptible, CDT- majority and a highly resistant, hypervirulent (CDT+) minority.

Integrating genomic location and sequence contexts of point mutations to improve classifying cancer tissue of origin

Anh Phuong Le (Phuong)[1], Cheng Soon Ong (Cheng)[2], Gavin Huttley[1]

1. *Research School of Biology, The Australian National University*
2. *Data61, CSIRO*

Both the genomic location of mutations along and the immediate sequence neighbourhood have proven informative for predicting cancer tissue of origin. These aspects are considered to reflect, respectively, the influence of chromatin state on mutagenesis and the influence of local sequence context on lesion formation and repair. Of these, the genomic location effect is conventionally represented as non-overlapping 1MB bins; the sequence context effect is conventionally represented as a strand-symmetric mutations occurring in strand asymmetric 3-mer neighbouring bases. In my honours, I tackled the basis for these representations by considering the phenomena from both a

statistical and machine learning perspectives. I used publicly available data from the International Cancer Genome Consortium and ENCODE. Statistical analyses established that chromatin accessibility was more variable between cancers than previously considered, and that the information in sequence context extended beyond 3-mers. On the 12 cancers employed in my study, the best classification algorithm combined a continuous representation of genomic location with a strand asymmetric trinucleotide representation of point mutations, as opposed to the standard representation.

Identification of the most influential nodes involving all topological dimensions of a network

Adrian (Abbas) Salavaty[1], Mirana Ramialison[1], Peter D Currie[1]

1. *Australian Regenerative Medicine Institute, Monash University, Clayton, VIC 3800, Australia*

Biological systems are composed of highly complex networks, and decoding the functional significance of individual network components is critical for understanding healthy and diseased states. Several algorithms have been designed to identify the most influential nodes within a network. However, current methods do not address all the topological dimensions of a network or correct for inherent positional biases, which limits their accuracy. Here we present Integrated Value of Influence (IVI), an algorithm that integrates the most important and commonly used network centrality measures in an unbiased way and captures all of the topological dimensions of a network to successfully identify the most influential nodes. The evaluation of IVI in the context of both simulated and experimental data confirmed their superiority to other respective contemporary methods and algorithms. Altogether, IVI is a versatile algorithm that could help all network researchers in the identification of the most influential players in the entire system.

An automated meta-caller to detect de novo mutations from whole genome trio data using cloud computing technology

Anushi Shah[1], Michael Troup[1], Eleni Giannoulatou[1]

1. *Victor Chang Cardiac Research Institute*

De novo mutations (DNMs) are genetic alterations occurring for the first time in a family member. DNMs have been shown to be a significant cause of severe developmental genetic disorders. Current estimates show that the number of de novo Single-Nucleotide Variants (SNVs) in the genome of an average individual ranges from 44 to 82. However, our analysis shows that current DNM callers still produce thousands of false-positive DNMs. Each DNM caller has unique pre-processing steps and resource requirements making processing large genomics data cumbersome and time consuming. By investigating the patterns of allelic fraction (AF) and read depth (DP) from whole genome sequencing(WGS) data we have observed that often the false positive DNMs exhibit similar properties to true positive DNMs. By using the consensus of four DNM callers (Varscan, DenovoGear, TrioDenovo and PhaseByTransmission), we were able to detect DNMs more accurately and filter out many false positives. Hence, we have developed an automated meta-caller to detect DNMs from trio WGS data. We harness elastic capabilities of cloud technology with an automated workflow engine (Cromwell) for efficient processing. The meta-caller produces 10 times less false positive variants maintaining sensitivity. It provides a list of concordant DNMs across all four DNM callers. The entire

process is executed in an automated fashion on cloud resources. Our DNM meta-caller is a useful tool for primary analysis in disease sequencing studies where DNMs could be the cause of disease. Automation enables the execution of the meta-caller processing pipeline in a faster, scalable and parallel fashion.

Landscape genetics to determine factors for ongoing transmission of onchocerciasis in the transition region of Ghana

Himal Shrestha[1], Shannon M Hedtke[1], Karen McCulloch[1,2], Warwick N Grant[1]

1. *Department of Physiology, Anatomy and Microbiology, School of Life Sciences, La Trobe University, Bundoora, Australia*
2. *2. WHO Collaborating Centre for Viral Hepatitis, Victorian Infectious Diseases Reference Laboratory, Royal Melbourne Hospital, and Department of Infectious Diseases, University of Melbourne, at the Peter Doherty Institute for Infection and Immunity, Australia*

Onchocerciasis is a neglected tropical disease caused by a filarial parasite, Onchocerca volvulus, and transmitted by the bites of black flies (Simulium spp.). Since the 1990s, onchocerciasis control has been done with the mass drug administration with ivermectin (MDAi) in endemic communities. While transmission has been reduced significantly in most of the endemic areas, there are a few persistent foci spread throughout different endemic countries. This can be a threat to neighbouring foci where the transmission has been successfully controlled, posing a risk of resurgence of the disease and wasting decades of intervention efforts. While estimates of relatedness using genetic data have revealed whether parasites have moved between locations, it is critical to determine the ecological factors that influence the connectivity between those sites and thus may allow ongoing movement. We present a landscape genetics approach to determine the ecological factors influencing the connectivity between sampling sites. We analysed mitochondrial SNP data from 163 worms from 14 communities in the transition region of Ghana characterised by persistent onchocerciasis transmission. We found environmental variables such as elevation and land cover were responsible for maintaining the connectivity between the sampling locations. We were able to identify the key areas in the transect of Ghana which are maintaining a high level of connectivity between the parasite populations. With the abundance of global remote sensing and genetic data, landscape genetics is a timely and powerful tool to address the problem of persistent foci for the global elimination of filarial diseases.

Genomic landscape of diversification, selective sweeps, and demographic history of an anthroponotic parasite

Swapnil Tichkule[1], Simone M. Cacciò[2], Guy Robinson[3], Rachel M. Chalmers[3], Samantha J. Emery-Corbin[1], Kevin M. Tyler[4], Cock van Oosterhout[5], Aaron R. Jex[1]

1. *Population Health and Immunity, Walter and Eliza Hall Institute of Medical Research, Melbourne, VIC, Australia*
2. *Department of Infectious Disease, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161, Rome, Italy*
3. *Cryptosporidium Reference Unit, Public Health Wales Microbiology, Singleton Hospital, Swansea, UK*

4. *Biomedical Research Centre, Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich, UK*
5. *School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich, UK*

Cryptosporidium is a significant public health problem and one of the primary causes of diarrhoea in humans, particularly in very young children living in low- and middle-income countries. The zoonotic Cryptosporidium parvum and anthroponotic C. hominis species account for most cases globally, but the latter is predominant in low- and middle-income countries. Here, we present a comprehensive whole genome study of C. hominis, comprising 114 isolates from 16 countries in five continents. We detect two highly diverged lineages with a distinct biology and demography that have diverged circa 500 years ago. We consider these lineages as two subspecies, and provisionally propose the names C. hominis hominis (clade 1) and C. hominis acquapotentis (clade 2). C. hominis hominis is mostly found in low-income countries in Africa and Asia, and it appears to have recently undergone population contraction and selective sweep. In marked contrast, we reveal a signature of population expansion in C. hominis acquapotentis found in high-income countries, mainly in Europe, North America, and Oceania. Moreover, we detect genomic regions of introgression representing gene flow after a secondary contact between the subspecies from low- and high-income countries. Furthermore, we show that this gene flow resulted into genomic island of high diversity and divergence, and we find that diversity at potential virulence genes is maintained by balancing selection, which suggests that they are involved in a coevolutionary arms race.

Using multimodal single cell data to predict regulatory gene relationships and to build a computational direct cell reprogramming model

Andy Tran[1], Pengyi Yang[1], Jean Y.H. Yang[1], John Ormerod[1]

1. *University of Sydney*

Cell reprogramming offers a potential treatment to many diseases, by regenerating specialized somatic cells. Despite decades of research, discovering the transcription factors that promote direct cell reprogramming has largely been accomplished through trial and error, a time-consuming and costly method. A computational model for direct cell reprogramming, however, could guide the hypothesis formulation and experimental validation, to efficiently utilize time and resources. Current methods often cannot account for the heterogeneity observed in reprogramming, or they only make short-term predictions, without modelling the entire reprogramming process. Here, we present scREMOTE, a novel computational model for direct cell reprogramming that leverages single cell multiomics data, enabling a more holistic view of the regulatory mechanisms at cellular resolution. This is achieved by first identifying the regulatory potential of each transcription factor and gene to uncover regulatory relationships, then a regression model is built to estimate the effect of transcription factor perturbations. We show that scREMOTE successfully predicts the long-term effect of overexpressing two key transcription factors in hair follicle development by capturing higher-order gene regulations. Together, this demonstrates that integrating the multimodal processes governing gene regulation creates a more accurate model for direct cell reprogramming with significant potential to accelerate research in regenerative medicine.

Three-minute flash talks

Listed in alphabetical order of presenters

Characterizing changes in gene co-expression facilitating the loss of features of multicellularity driving prostate cancer progression

Mikhail Dias[1], Dr David Goode[1], Dr Anna Trigos[1]

1. *Computational Biology Program, Peter MacCallum Cancer Centre, Sir Peter MacCallum Department of Oncology, The University of Melbourne*

The transition to multicellularity involved evolution of gene regulatory networks (GRN) to coordinate and maintain cellular processes in order to promote organism-level fitness. Transcriptomic analysis of data from The Cancer Genome Atlas has revealed networks acquired during the transition to multicellularity are often broken down in cancer leading to tumorigenesis. We aim to uncover how these pathways are rewired in Prostate cancer (PC) to evade treatment. We have developed Evolutionary Network Analysis (ENA) a unique multi-omics approach combining evolutionary analysis, transcriptomics and network biology to investigate how GRNs acquired during the transition to multicellularity are rewired in cancer. Applying ENA to PC patient samples stratified by progression of benign to malignant and primary to metastatic tumours, created a comprehensive landscape of changes in gene co-expression during PC progression. This enables investigations of how tumours are able to rewire GRNs that evolved to support multicellularity to access pathways that facilitate tumour progression and acquired drug resistance. Our analysis reveals gene co-expression modules become progressively more disrupted and rewired as PC advances to higher Gleason grade groups. Additionally, preliminary results of co-expression modules indicate genes acquired during the transition to multicellularity are progressively rewired as the Gleason grade group increases suggesting that these new connections enable tumours to activate more ancient unicellular pathways to survive and progress. This project demonstrates how utilizing gene co-expression signatures can be used to gain a comprehensive molecular landscape of PC, which is immensely valuable for the development of more robust therapeutic strategies.

Molecular docking between herbal formula compounds and stroke-related numbness and weakness targets

Sanghyun Lee[1], Ann Rann Wong[1], Angela Wei Hong Yang[1], Andrew Hung[2]

1. *School of Health and Biomedical Sciences, RMIT University, Bundoora, Victoria 3083, Australia*
2. *School of Science, RMIT University, Melbourne, Victoria 3000, Australia*

Stroke related numbness and weakness (SRNW) are resultant disabilities following a stroke, and 85% of post-stroke patients suffer from it. Huangqi Guizhi Wuwu Tang (HGWT), consisting of five herbs, is a traditional herbal formula used for managing post-stroke symptoms. This study aimed to investigate the mechanisms of HGWT targeting SRNW-related proteins using a molecular docking method. SRNW-related protein targets were obtained from DrugBank and Open Target Databases. Chemical compounds of HGWT ingredients were identified from TCMSP and PubChem. Computational molecular docking between targets and compounds was performed using the PyRx

and AutoDock Vina. A total of 14,858 docking results were obtained between 19 targets and 782 compounds. The binding affinity scores ranged from -0.8 to -11.1 kcal/mol and their average binding score was -6.0 kcal/mol. The highest binding affinity was between DZ118 (Fumarine) and ESR1 (Estrogen receptor alpha) with a binding affinity score of -11.1 kcal/mol. The second strongest binding affinity was between PDE5A (cGMP-specific 3',5'-cyclic phosphodiesterase) and BS040 ((3S,5R,8R,9R,10S,14S)-3,17-dihydroxy-4,4,8,10,14-pentamethyl-2,3,5,6,7,9-hexahydro-1H-cyclopent a[a] phenanthrene-15,16-dione) with a binding affinity score of -10.7 kcal/mol. The third highest binding affinity (-10.4 kcal/mol) was predicted for two ligand-protein complexes, including HQ068 (Asernestioside B) against MAPK3 Mitogen-activated protein kinase 3) and DZ006 (Spiradine A) against PDE5A. Further research is required for determining the stability of the predicted ligand-receptor complexes using more rigorous computational simulation methods.

A bioinformatics workflow and Shiny app for 16S microbiota community analysis and data visualisation in Acute Lymphoblastic Leukaemia

Holly Martin[1], Feargal Ryan[1], Melanie Smith[1,2], Jacqueline Rehn[1], Stephanie Field[1], Kiralee Vuglar[1], Deborah White[1,3], Jimmy Breen[1]

1. *The University of Adelaide, South Australian Health and Medical Research Institute*
2. *Flinders University, South Australian Health and Medical Research Institute*
3. *The University of Adelaide, University of South Australia, South Australian Health and Medical Research Institute*

The human gut microbiota is an emerging target for clinical intervention to improve patient outcomes, treatment responses and overall health. Data visualisation is an important consideration in communicating results in an accurate and appealing way to researchers and clinicians. In this project we build a bioinformatics workflow and visualisation tool to aid analysis of gut microbiota 16S rRNA gene sequencing data from Acute Lymphoblastic Leukaemia (ALL) patients. We developed 'shinyMicrobiota', an R/Shiny app for interactive visualisation and exploration of this data (https://github.com/hollyanitamartin/shinyMicrobiota). shinyMicrobiota uses outputs from the workflow to produce plots displaying taxonomic composition, alpha diversity, beta diversity and bacteroidetes to firmicutes ratios. These plots can be customised based on variables in the uploaded metadata and includes features such as data subsetting, grouping by variables, choice of plot type, statistical testing approach, plot title and choice of colour palettes. A standardised bioinformatics and visualisation workflow was used to analyse three independent datasets; a preclinical mouse model of ALL, a clinical dataset from ALL patients and a pre-diabetes public dataset from the Human Microbiome Project. These analyses demonstrate the functionality of the workflow and Shiny app across various experimental designs (animal model, clinical data and public dataset) and scales (34-398 samples). This demonstrates the ability of the workflow and shinyMicrobiota to scale to larger datasets and across both preclinical and clinical data.

Getting a foothold in Pangenomics

Chelsea Matthews[1], Nathan Watson-Haigh[2]

1. *School of Agriculture, Food and Wine and ARC Centre of Excellence in Plant Energy Biology, University of Adelaide, Waite Campus, Urrbrae, SA 5064, Australia*
2. *South Australian Genomics Centre, SAHMRI, North Terrace, Adelaide SA 5000, Australia*

Pangenomic models are composed of genomic data from multiple members of a single species. They focus on identifying genomic sequence/genes present in all members of the species - termed the core genome - and genomic sequence/genes present in only some members of the species - termed the accessory genome. Much research centers around using pangenomic models in the place of linear reference genomes as they have been shown to increase the accuracy of variant calling and are well suited to linking phenotype with genotype. However, the language used around pangenomics is ambiguous with little delineation between different types of pangenomes. This makes it difficult for people new to the field to gain a foothold. Here we explore the different types of pangenomic models and furnish them with descriptive names. Hopefully these names can be used in discussions around pangenomics and will help to clear up any potential confusion for people new to the field.

BITHub: An interactive web resource of gene expression in the human brain

Urwah Nawaz[1]*, Kieran Walsh[2]*, Gavin Sutton[2], Jozef Gecz[1], Irina Voineagu[2]

**Equal contribution*

1. *Adelaide Medical School, Robinson Research Institute, University of Adelaide, SA, Australia*
2. *School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia*

Large-scale transcriptomic studies have been instrumental in characterising gene expression in the human brain across brain regions, developmental stages and disease states. Several large-scale consortia have generated gene expression data from the human brain including GTEx, PsychENCODE, CommonMind, BrainSeq and BrainSpan. The data generated by these consortia partially overlaps in terms of sample characteristics, and has been generated with different library preparation and data analysis methods. Despite this wealth of data, it is currently difficult to extract and compare gene expression information across these large-scale datasets. Here, we present the Brain Integrative Transcriptome Hub (BITHub), a web resource that aggregates gene-expression data from the human brain across multiple consortia, and allows direct comparison of gene expression in an interactive manner. BITHub also allows the exploration of co-variates that drive gene expression variance in each dataset. We believe that this user-friendly web resource will be valuable for neuroscience and neurogenetics research by allowing biologists and clinical geneticists to uncover patterns of gene expression in the human brain, and to determine whether observations are replicable across datasets.

In silico studies to examine the interaction of phytochemicals from a medicinal herb Wolfiporia cocos with peroxisome proliferator-activated receptor delta (PPARδ)

Ann Rann Wong[1], Andrew Hung[2], Angela Wei Hong Yang[1], Harsharn Gill[2], George Binh Lenon[1]

1. *School of Health and Biomedical Science, RMIT University, Melbourne, Victoria, Australia*
2. *School of Science, RMIT University, Melbourne, Victoria, Australia*

Wolfiporia cocos (WC), also known as Fu ling, is a medicinal herb frequently included in Chinese herbal formulations for the management of obesity. Improvements in body weight, body fat percentage and blood lipid profiles have been reported among WC-containing formulations in clinical trials. However, studies on the mechanisms by which WC ligands target obesity receptors remain limited. Therefore, this study aimed to examine the interaction of WC ligands with PPARδ, a regulator of glucose and lipid metabolism, mostly expressed in skeletal muscle and adipose tissue. Forty-four WC ligands were retrieved from TCMSP and TCM-ID databases and docked with PPARδ (PDB:5U46) based on AutoDock Vina default parameters. The binding affinities of 44 ligands ranged from -4.7 to -10 kcal/mol with a mean and standard deviation of -7.0±1.7 kcal/mol. Unlike the known agonist GW501516, most of the WC ligands preferentially bind to a region proximal to the β-sheets and ω-loop of the relatively large ligand-binding domain. As WC ligands form contacts with differing residues to known agonists, it may be possible that WC ligands could initiate allosteric effects. Among the 16 hit ligands with a binding affinity of ≤ -8 kcal/mol, 93.8% complied with 4 or all of Lipinski's rule, 100% complied with Veber's rule, and two ligands (WC_TCMC878 and WC_TCMC24) were predicted to be potential leads. Overall, this study identified several high-binding WC ligands which may modulate PPARδ transcriptional activities in energy homeostasis. Further computational and experimental studies are underway to validate the anti-obesity potentials of WC ligands.

Potential non-coding regulations in cSCC Metastasis cancer

Amarinder S. Thind[1,2], Bruce Ashford[1,2,3,9], Dario Strbenac[4], Ruta Gupta[5], Jonathan R Clark[4,6,7], N. Gopalakrishna Iyer[8], Jenny Mitchell[3], Jenny Lee[9,10], Simon A Mueller[7,11], Elahe Minaei[2,12], Jay Perry[2,12], Marie Ranson[2,12]

1. *School of Medicine, University of Wollongong, Wollongong, NSW, Australia*
2. *Illawarra Health and Medical Research Institute, Wollongong, NSW, Australia*
3. *Illawarra Shoalhaven Local Health District Wollongong, NSW, Australia*
4. *Sydney Medical School, Faculty of Medicine and Health, The University of Sydney, NSW, Australia*
5. *Anatomical Pathology, Royal Prince Alfred Hospital, Sydney, NSW, Australia.*
6. *Royal Prince Alfred Institute of Academic Surgery, Sydney Local Health District, Sydney, NSW, Australia*
7. *Department of Head and Neck Surgery, Chris O'Brien Lifehouse, Sydney, NSW, Australia*
8. *National Cancer Center, Singapore*
9. *Department of Medical Oncology, Chris O'Brien Lifehouse, Sydney, NSW, Australia*
10. *Department of Clinical Medicine, Macquarie University, Sydney, NSW, Australia*
11. *Department of Otorhinolaryngology, Head and Neck Surgery, Zurich University Hospital and University of Zurich, Zurich, Switzerland*
12. *School of Chemistry and Molecular Bioscience, University of Wollongong, Wollongong, NSW, Australia*

There is limited published data exploring whole-genome sequencing (WGS) in metastatic cutaneous squamous cell carcinoma (cSCC). We used WGS on matched tumour and blood DNA to detect somatic variants from 25 patients with regional metastases of head and neck cSCC. Computational analyses at coding and non-coding levels are performed utilising a combination of bioinformatic tools to interrogate their clinical impacts on metastasis across the cohort. miRNA binding locations in 3'UTR were significantly functionally altered in EVC (71%), PPP1R1A (71%) and LUM (24%). Recurrent variation was observed in the tumour suppressing lncRNA LINC01003 in 68% of specimens. Recurrent copy number loss in tumour suppressor genes KANSL1 and PTPRD and gain in CALR, CCND1 and FGF3 was observed. Single nucleotide variation with the greatest functional impacts was most frequently observed in TP53, CDKN2A, ZNF750, OR51S1 and TET2. Metastatic cSCC is characterized by a highly mutated genome, most pronounced in the non-coding region, with recurrent patterns of variation in key regulatory elements, and recurrent copy number and short variation in cancer-associated genes.

Comprehensive transcriptional analysis of the CD4+ memory compartments in Tconv & Treg

Holly A. Withers[1], Chris M. Hope[2], Cheryl Y. Brown[2], Simon C. Barry[2]

1. *School of Biological Sciences, University of Adelaide, SA, Australia*
2. *Molecular Immunology, Department of Paediatrics, University of Adelaide, SA, Australia*

Immunological memory is a critical feature of the adaptive immune system. Durable memory responses are necessary to mount sustained, secondary defenses against pathogens and is the basis for vaccination. However, dysregulation of immunological memory can drive poor disease outcomes of chronic viral infection, such as COVID-19.

CD4+ T cells are central to all adaptive immune responses as they govern the functions of surrounding cells. They are a diverse population of cells comprising proinflammatory Tconv and FOXP3+ suppressive Treg. Gene expression in Tconv and Treg must be stringently regulated to achieve immune homeostasis.

Tconv and Treg primarily exist in four memory stages: Naïve (TN), Central Memory (TCM), Effector Memory (TEM) and CD45RA+ Effector Memory (TemRA). TemRA are a relatively rare, ill-defined subtype, described as being terminally-differentiated and functionally heterogeneous. However, recent literature highlights a significant role of Tconv TemRA in viral clearance and tumour cytotoxicity. The Treg memory compartments remain uncharacterised.

We performed bulk RNA-seq in Tconv and Treg memory subsets from buffy coats of healthy individuals (N=5). Trait association analysis (via WGCNA) highlighted distinct transcriptional signatures, which effectively delineated all cellular phenotypes. Interestingly, we observed significant enrichment of a NK-cell gene signature in Tconv and Treg TemRA, compared with other cell types. GSEA between Tconv and Treg TemRA revealed enrichment of a Treg TemRA signature in Chronic lymphocytic leukemia patients, whilst Tconv TemRA genes were enriched in healthy patients, suggesting a possible cytolytic role of Tconv TemRA against neoplastic cells.

Currently, the Tconv and Treg memory compartments are being isolated from the peripheral blood of convalescent COVID-19 patients for bulk RNA-seq to investigate whether COVID-19 infection leads to dysregulation of memory T cell gene signatures, particularly in TemRA.

Generating personalised tumour specific PCR primers for recurrence monitoring by liquid biopsy utilising structural variants identified from 30x FFPE tumour tissue WGS data without matched normal samples

Boris Ka Leong Wong[1], Gavin Wright[2], Stephen Barnett[3], Hongdo Do[4], Alexander Dobrovic[1], Daniel Cameron[5]

1. *Translational Genomics and Epigenomics Laboratory, Department of Surgery, The University of Melbourne, Australia*
2. *Department of Surgery, St. Vincent's Hospital Melbourne, The University of Melbourne, Fitzroy VIC, Australia*
3. *Department of Cardiothoracic Surgery, Austin Hospital, Heidelberg VIC, Australia*
4. *Pathology Department, St Vincent's Hospital Melbourne, Fitzroy VIC, Australia*
5. *Bioinformatics Division, Walter and Eliza Hall Institute of Medical Research, Parkville VIC, Australia*

Aim

To create a bioinformatics tool that can:

1. Filter confident somatic structural variants (SVs) from 30x formalin-fixed, paraffin-embedded (FFPE) tumour tissue whole genome sequencing (WGS) data without matched normal samples
2. Generate personalised tumour specific PCR primers targeting the filtered SVs for longitudinal recurrence monitoring by measuring plasma circulating tumour DNA (ctDNA) levels

Background

DNA fragmentation and sequence artifacts commonly occur in DNA extracted from routine pathology FFPE tissues, which lead to high false positive SV calling rate. It is also challenging to detect somatic SVs from 30x WGS data without germline subtraction from normal tissue WGS data. Furthermore, Designing PCR primers targeting multiple SVs in a large group of patients for ctDNA detection can be tedious.

Therefore, a tool that can select confident somatic SVs results and generate PCR primers for ctDNA detection is useful in routine pathology practice.

Method

DNA was extracted from 54 FFPE tumours followed by 30x WGS. SVs were identified by GRIDSS and filtered by GRIPSS. The SV results were filtered by our tool. The remaining SVs were fed to primer3 for primer design. The output primers were piped to blat for quality check. The top 5 primers with highest GRIDSS were validated with droplet-digital PCR on blank control, tumour and germline DNA.

Conclusion

>4500 SVs were identified after GRIPSS filtering. <30 SVs remained after applying our tool. The PCR validation result indicated that at least 1 somatic primer set was generated for 49/54 (91%) of tumours with 3 somatic primers for each tumour on average.