COMBINE

Australian Computational Biology and Bioinformatics Student Society

ISCB REGIONAL Student GROUP Australia

Bioinformatics Student Symposium

In-person conference
4th December 2023
Brisbane

COMBINE 2023 Symposium

Welcome

Welcome to the 2023 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 throughout the symposium.

We hope you enjoy the day.

Regards,
The 2023 COMBINE Symposium Committee

2023 COMBINE Symposium Committee

Chairs
Sam Davis & Oliver Hughes

Committee Members
Aliah Aziz, Lachlan Baer, Shane Bao, George Bouras, Sehaj Dhariwal, Alecia Gee, Eileen Ho, Laura Inglis, Hisatake Ishida, Bhavika Kumar, Namuhan, Sumedha Negi, Sebastian Porras, Sophie Shen, Vera Sun, Jenna Supper & Zoe Wu

COMBINE 2023 Symposium

Acknowledgement of Country

We acknowledge the land that we meet on for the COMBINE 2023 Symposium is the traditional lands for the Turrbal people of the north side of the river and the Jagera people of the south side of the river and we respect their spiritual relationship with their country.

We also acknowledge the Turrbal people and the Jagera people as the Traditional Custodians of Meanjin (Brisbane) and we pay our respects to Elders past, present and emerging.

We also pay respects to the cultural authority of Aboriginal people visiting/attending from other areas of Australia.

COMBINE 2023 Symposium

Symposium information

Venue

Address
Hilton Brisbane, 190 Elizabeth St, Brisbane City QLD 4000, Australia

Public transport
Hilton Brisbane is conveniently located near multiple public transport options. Queen Street bus station is a 5 minute walk, and Central Street train station is about a 10 minute walk. There are also multiple bus stops located along Elizabeth Street.

Parking
There is limited valet car parking available at the venue at the cost of $58 per vehicle per day and is subject to availability. The hotel does not offer discounted car parking or validation.

Links

Conference website: https://2023.abacbs.org/
COMBINE website: https://www.combine.org.au/
COMBINE Twitter: https://twitter.com/combine_au
COMBINE slack: https://combine-au.slack.com/
ABACBS Twitter: https://twitter.com/abacbs
ABACBS website: https://www.abacbs.org/
Hilton Brisbane: https://www.hilton.com/en/hotels/bsbhitw-hilton-brisbane/
Queensland public transport: https://translink.com.au/

COMBINE 2023 Symposium

Program overview

The times indicated in the following program are in Australian Eastern Standard Time (Brisbane, GMT+10).

08:00 – 08:30	**Registration**
08:30 – 08:45	**Welcome**
08:45 – 09:30	**Session 1: Biomedical Informatics**
09:30 – 10:30	**Poster Session**
10:30 – 11:00	**Morning Tea**
11:00 – 11:45	**Keynote Address**
11:45 – 12:30	**Session 2: Microbial Genomics**
12:30 – 13:15	**Lunch**
13:15 – 14:15	**Session 3: Machine Learning**
14:15 – 14:30	**Break**
14:30 – 15:15	**Session 4: ‘Omics Methodology**
15:15 – 15:45	**Afternoon Tea**
15:45 – 16:30	**Careers Panel**
16:30 – 17:00	**Closing**

COMBINE 2023 Symposium

Invited speakers

Keynote speaker

Prof. Jean Yang

School of Mathematics and Statistics and
Director of Sydney Precision Data Science Centre,
The University of Sydney

Professor Yang is a leader in statistical bioinformatics who works in cutting-edge biomedical research. She is currently the Director of the Sydney Precision Data Science which she established in 2022. Her research has led to seminal advances in integration of multi-layered biological data integration by removing extraneous variability and accounting for heterogeneity. Recently, she has made similarly significant advances in scalable data integration in single-cell transcriptomics sequencing. She was awarded the Moran Medal in recognition of her work on developing methods for molecular data arising in cutting-edge biomedical research. As a statistical data scientist who works in the interface of statistics, biomedicine, and health. She enjoys developing novel methods with translational potential in a collaborative environment, working closely with investigators from diverse backgrounds.

COMBINE 2023 Symposium

Invited speakers

Careers panel

Dr Anne Klein - Postdoctoral Research Fellow, Digital Genome Engineering Team, CSIRO

In 2012, Anne completed a bachelor's degree in Biomedical Sciences, followed by a Master's in Immunology, and subsequently another Master's in Bioinformatics from the University Paris Diderot. Anne pursued her PhD at the University of the Sunshine Coast in 2018, delving into the omics landscape of invertebrates. Upon completing her PhD, she secured a post-doctoral fellowship with CSIRO, specifically within the Transformational Bioinformatics team. There, she worked on Genome Engineering under the guidance of Dr. Laurence Wilson. Currently, in the scope of gene therapy, her research mainly focuses on optimizing the packaging efficiency of viral vectors using in silico methods and developing user-friendly bioinformatics tools.

Dr Olga Kondrashova - Team Lead, QIMR Berghofer

Olga is a NHRMC Emerging Leadership Fellow and leads the Molecular Oncology team at the QIMR Berghofer Medical Research Institute, Brisbane. Olga's team utilises bioinformatic and machine learning approaches to link cancer genomics with patient treatment responses. Olga obtained her PhD at the University of Melbourne in 2016 on molecular profiling of ovarian cancer. She then conducted her postdoctoral studies at WEHI, where her research led to the identification of two novel mechanisms of sensitivity and resistance to PARP inhibitors in ovarian cancer.

Dr James McCafferty - Chief Information Officer at Wellcome Sanger Institute

James is the Chief Information Officer at the Wellcome Sanger Institute. His role covers IT strategy, delivery and operations to support the goals of the Institute. This encompasses research IT, research data, research software and informatics, enterprise applications, IT infrastructure and information governance and security. James has over 30 years experience in IT systems and networks. Before joining the Sanger, James was Chief Information Officer and Director of Research IT at University College London. And before that, various IT-related and programme management roles in BT (the UK's leading telecommunications provider). James' PhD is in computer vision. James' current key interests centre around ensuring that the Sanger's Informatics and Digital Solutions play a very active role in helping scientists plan and deliver bold and ambitious science.

Dr Stevie Pederson - Postdoctoral Bioinformatician, Telethon Kids Institute

Dr Pederson completed their PhD at the University of Adelaide in 2018, whilst simultaneously working as the Coordinator of the Bioinformatics Hub. They have a long history in statistics, bulk transcriptomics and R package development, being senior author for 4 Bioconductor packages and making contributions to multiple additional packages. In 2020, Dr Pederson joined the Dame Roma Mitchell Cancer Research Laboratories, gaining deep expertise in ChIP-Seq data before moving to Black Ochre Data Labs in 2022, within the Indigenous Genomics unit at Telethon Kids Institute. Their current area of research focus is on a large cohort of bulk transcriptomes, addressing risk factors for Type 2 Diabetes and subsequent complications within the Indigenous Australian community and extending to the integration across multiple omics layers. Dr Pederson has recently joined the Bioconductor Community Advisory Board which endeavours to ensure equitable access to the resources of Bioconductor both locally and internationally.

COMBINE 2023 Symposium

Detailed program

Registration
08:00 – 08:30 (Foyer)

Symposium Welcome
08:30 – 08:45 (Ballroom A)

AEST	***Chairs:*** *Sam Davis, Oliver Hughes*
08:30 – 08:45	**Acknowledgement of Country** **Welcome to COMBINE 2023**

Session 1: Biomedical Informatics
08:45 – 09:30 (Ballroom A)

AEST	***Chair:*** *Sam Davis*
08:45 – 09:00	**CaraVaN: Prioritising pathogenic cardiac variants in the non-coding genome with ensembling** Ngoc Duy Tran
09:00 – 09:15	**Methylation profile scoring model for discovering epigenetic evidence on endometriosis** Li Ying Thong
09:15 – 09:30	**GABAA ion channels as therapeutic targets in Glioblastoma** Ishika Mahajan

Poster Session
09:30 – 10:30 (Queen's Ballroom)

Morning Tea
10:30 – 11:00 (Foyer)

Keynote Address
11:00 – 11:45 (Ballroom A)

AEST	***Chair:*** *Aliah Aziz*
11:00 – 11:45	**Prof. Jean Yang** School of Mathematics and Statistics and Director of Sydney Precision Data Science Centre, The University of Sydney

Session 2: Microbial Genomics
11:45 – 12:30 (Ballroom A)

AEST	***Chair:*** *George Bouras*
11:45 – 12:00	**Microbiomes and machine learning: large scale prediction of environmental conditions using taxonomic profiles** Joshua Mitchell
12:00 – 12:15	**Genomic insights into the longitudinal transmission of Neisseria gonorrhoeae in Victoria** Mona Taouk
12:15 – 12:30	**Genome-resolved metagenomics reveal novel viral-host associations in a subtropical estuary** Apoorva Prabhu

Lunch
12:30 – 13:15 (Foyer)

Session 3: Machine Learning
13:15 – 14:15 (Ballroom A)

AEST	***Chair:*** *Lachlan Baer*
13:15 – 13:30	**PRIMITI: a machine learning model for precise identification of novel miRNA-target mRNA repression** Korawich Uthayopas
13:30 – 13:45	**Deciphering the Black Box: Mastering the MSA Transformer for Phylogenetic Tree Reconstruction** Ruyi Chen
13:45 – 14:00	**DDMut-PPI: predicting effects of mutations on protein-protein interactions using deep learning** Yunzhuo Zhou
14:00 – 14:15	**Overcoming Heterogeneity for Improved CNN-Based Classification of Replicative Senescence** Ebony Watson

Break
14:15 – 14:30

Session 4: ‘Omics Methodology
14:30 – 15:15 (Ballroom A)

AEST	***Chair:*** *Oliver Hughes*

14:30 – 14:45	**Faster off-target risk assessment for CRISPR guide RNAs** Carl Schmitz
14:45 – 15:00	**Damsel: an R package for end to end analysis of DamID** Caitlin Page
15:00 – 15:15	**Improving Spatial Domain Clustering and Trajectory Inference with Dynamic Biologically-informed Graph Structure and Graph Convolutional Neural Network** Roxana Zahedi Nasabi

Afternoon Tea
15:15 – 15:45 (Foyer)

Careers Panel
15:45 – 16:30 (Ballroom A)

AEST ***Chairs:*** *George Bouras, Laura Inglis*

15:45 – 16:30

Dr Anne Klein
Postdoctoral Research Fellow, Digital Genome Engineering Team, CSIRO

Dr Olga Kondrashova
Team Lead, QIMR Berghofer

Dr James McCafferty
Chief Information Officer at Wellcome Sanger Institute

Dr Stevie Pederson
Postdoctoral Bioinformatician, Telethon Kids Institute

Symposium Closing
16:30 – 17:00 (Ballroom A)

AEST ***Chairs:*** *Sam Davis, Oliver Hughes*

16:30 – 17:00

Prize announcements

President's report

COMBINE 2023 Symposium

Speaker abstracts

Biomedical Informatics

Ngoc Duy Tran, The Murdoch Children's Research Institute

CaraVaN: Prioritising Pathogenic cardiac variants in the non-coding genome with ensembling

Congenital Heart Diseases (CHD) is the most prevalent birth defect due to errors in fetal heart development, resulting in approximately 1% of newborns affected worldwide annually. Although whole genome sequencing (WGS) has enabled detecting variants in CHD patients, the function for many of those remains unknown as they are detected in the non-coding part of the genome where our understanding is still limited.
This work presents CaraVaN, a cardiac-specific supervised learning model specialized in annotating and identifying potential CHD pathogenic variants in the non-coding region. This model learns from cardiac-specific consequence features: candidate cREs from ENCODE, EpiMap and zebrafish cardiomyocytes (n=74), transcription factor binding sites (n=98), distal chromatin interactions using HiC technique (n=3), ATAC-seq from ENCODE (n=34) and prediction scores from existing variant assessment tools (n=2). By extracting useful features, CaraVaN utilizes ensembling method which incorporates Logistic Regressions with 2 gradient boosting methods namely LGBM and CatBoost to generate probabilistic predictions indicating the potential cardiac pathogenicity in CHD for each non-coding variant.
Regarding evaluation performance, CaraVaN has achieved highly competitive results (ROC AUC=0.865) compared to the state-of-art tools that are not tissue-specific (ROC AUC=0.689). CaraVaN is validated to prioritise a functionally known non-coding variants in CHD in chromosome 12. Gene ontology (GO) analysis on the top-scoring non-coding variants revealed their association with 57 genes involved in human heart disease phenotypes including atrial fibrillation, supraventricular arrhythmia and abnormality of cardiovascular system physiology. Overall, CaraVaN is the first tool to identify non-coding variants associated with CHD and other heart-related diseases.

Li Ying Thong, The University of Queensland

Methylation profile scoring model for discovering epigenetic evidence on endometriosis

Endometriosis is a disease where tissue lining the inside of the uterine cavity is found outside of the uterus. Studies have shown that it takes 4-11 years for a patient to receive a proper diagnosis after experiencing symptoms. DNA methylation (DNAm) can be influenced by both genetic and environmental factors, making it a useful biological marker to capture the effects of both genetic and environmental exposures on complex diseases. However, due to the difficulty in obtaining endometrial samples and limited sample size available, true associations between methylation sites and endometriosis is hard to determine. Thus, the use of methylation profile score (MPS), a more highly powered method to detect differentially methylated sites for endometriosis is explored. This study aims to develop and evaluate MPS for endometriosis using DNAm data from blood and endometrium tissue samples collected from 291 (84 controls; 207 surgically diagnosed cases) and 951 (341 controls; 610 surgically diagnosed cases) participants respectively. The performance of our MPS models from blood and endometrium samples in predicting endometriosis have achieved a maximum AUC of 0.5883 and 0.6002 respectively suggesting that there is a difference in DNAm profiles between endometriosis cases and controls. Furthermore, blood and endometrium samples have similar AUC indicating that DNAm from blood, a more accessible tissue, could be equally informative in providing insight on the epigenetic architecture of endometriosis. The future focus of this study is to improve the accuracy of our current MPS model in predicting endometriosis by incorporating genotype data and other endometriosis

risk factors into the model to extract information from both genetic and non-genetic factors that increase risk of endometriosis. Our work aims to provide insights on the epigenetic pathogenesis of the disease that could help shape a more effective diagnosis and treatment strategies.

Ishika Mahajan, The University of South Australia

GABAA ion channels as therapeutic targets in Glioblastoma

While the survival rates of most cancers have dramatically improved in the last few decades, this is not the case for brain cancers. For glioblastoma, the most diagnosed malignant brain cancer in adults, the statistics are far worse, with a 5- year survival of just 5%. This is due to the limited understanding of how tumour cells invade the surrounding healthy brain and the tumour microenvironment's contribution to it. Recently, we developed a bioinformatics pipeline utilizing AI, bulk-sequencing, single-cell, and spatial transcriptomics of glioblastoma tumour biopsies to identify new targets for invasive brain cancers. Our research indicates that GABAA ion channels (GABRG2 and GABRB1) expression occur primarily in tumour cells at the invasive front (Leading edge) and is highly correlated with patients' survival. Further, Through Drug Target analysis (bioinformatics and ML pipeline), we found FDA approved drugs targeting these channels and are currently used for other neurological disorders and cancers, but they are not yet used for glioblastoma patients. Our preliminary results using patient-derived explant organoids (PD-TEO) show that activating these channels through drug repurposing attenuate proliferation and reduce tumour growth. PD-TEOs are derived from the patient’s resected tumour tissue and therefore mimic and replicate the pathological, genetic, and treatment response tumour characteristics observed in patients, which other models fail to do. Further experiment validations can help us identify new therapies targeting these channels which can potentially lead to new clinical trials which, if successful, will improve the survival of glioblastoma patients and help us treat highly invasive brain tumours.

COMBINE 2023 Symposium

Speaker abstracts

Microbial Genomics

Joshua Mitchell, Queensland University of Technology

Microbiomes and machine learning: large scale prediction of environmental conditions using taxonomic profiles

Microbial life is everywhere, influencing and being influenced by local environs. These actions and reactions sum up, along with macroscopic and abiotic processes, to create the emergent properties of the world's ecosystems. As anthropogenic climate change intensifies, the fundamental biogeochemical characteristics of these biomes are being altered, such as the pH, temperature, and concentration of molecular oxygen. Therefore, expanding our understanding of how microbiota relate to these environmental properties will allow for a better prediction of how climate change may progress in future. How environmental factors influence community composition and vice versa is an understudied aspect of microbial ecology, in part due to the incredible diversity of microbial species, their functions, and their habitats. Here, we take, a big data approach to this problem, annotating all publicly available metagenome sequence datasets with global climate and geochemical estimates, creating a compendium of microbial communities annotated with their chemical and physical conditions (~40k with temperature and ~30k with pH so far, non-host associated). We then trained a number of machine learning algorithms to predict these factors in a wide variety of biomes using microbial community composition as input. These models can be used to retroactively annotate public microbiome data to improve their utility, discover novel information regarding biogeochemical cycles, infer the effects of environment on individual gene functioning and potentially improve global climate models.

Mona Taouk, The University of Melbourne

Genomic insights into the longtitudinal transmission of Neisseria gonorrhoeae in Victoria

The ongoing transmission of Neisseria gonorrhoeae, the cause of gonorrhoea, is a significant public health challenge. Despite improved access to prevention and treatment, Australia has seen a resurgence in gonorrhoea cases, with 33,692 reported in 2022, a 105.7% increase since 2012. We investigated the genetic diversity and antimicrobial resistance of N. gonorrhoeae isolates in Victoria from January 2017 to July 2021, including the COVID-19 period. Using cgMLST and Bayesian methods, we grouped 5,881 isolates into 233 genomic clusters, with 38.7% (12/31) of large clusters (≥30 genomes) showing persistent transmission for over 24 months. Persistence was significantly associated with larger clusters (OR 1.02, p=0.028) and clusters with a higher proportion of women (OR 2.91, P = 0.017). For the front-line dual treatments, phenotypic susceptibility to azithromycin was significantly associated with persistence (OR 127, p < 0.0001), while phenotypic resistance to ceftriaxone showed a near significant association with persistence (7.7, p = 0.057). Notably, the study explores the impact of COVID-19-related non-pharmaceutical interventions on N. gonorrhoeae transmission, indicating a decline in genomic diversity during this period. Bayesian phylodynamic techniques shed light on changes in the effective reproductive number before and after the pandemic measures.

The whole genome sequencing of N. gonorrhoeae isolates collected in Victoria has provided valuable insights into the longitudinal transmission patterns and genetic diversity of the pathogen. Phenotypic resistance to frontline treatment options reveals significant associations with cluster persistence,

underscoring the importance of ongoing monitoring of antimicrobial resistance and transmission dynamics of N. gonorrhoeae to inform effective control strategies and optimise treatment guidelines.

Apoorva Prabhu, The University of Queensland

Genome-resolved metagenomics reveal novel viral-host associations in a subtropical estuary

Tropical and subtropical estuaries represent dynamic ecosystems that serve as critical interfaces connecting terrestrial, marine, and freshwater biomes. They are considered hot spots for cycling of carbon derived from organic matter breakdown, and of nutrients that promote gross primary productivity. Microbial communities play a crucial role in estuary ecosystems by driving the biogeochemical processes, which affects marine food webs and ecosystem function. Viruses are the most enigmatic and abundant populations in estuaries, primarily regulating bacterial and archaeal mortality and also reprogramming host metabolic pathways through infection and lysis. Despite their ecological significance, viruses within subtropical estuaries have remained relatively understudied compared to their counterparts in the global oceans. In our work, we carried out a comprehensive characterization of the diverse viral populations in the subtropical Brisbane River Estuary and were able to link over 7,000 viruses to their microbial hosts. Employing genome-resolved metagenomics, we identified novel bacteriophages and archaeaviruses and assessed their host specificity. We inferred auxiliary genes, which may provide adaptive advantages for viral hosts under certain environmental conditions , and associated them with metabolic processes, such as photosynthesis, carbon, phosphorus and stress response. This pioneering study represents the first genomic exploration of viral communities and their potential ecological roles, in a subtropical estuary in the southern hemisphere.

COMBINE 2023 Symposium

Speaker abstracts

Machine Learning

Korawich Uthayopas, The University of Queensland

PRIMITI: a machine learning model for precise identification of novel miRNA-target mRNA repression

MicroRNAs (miRNAs) are a class of short non-coding RNAs that play a crucial role in the post-transcriptional regulation of key genes involved in various cellular pathways. Dysregulation of miRNAs has been implicated in the pathogenesis of multiple diseases, leading to the development of miRNA-based therapeutic applications such as biomarkers and drug targets. However, the research progress has been hindered by the limited understanding of miRNA repression mechanisms. To address the issue, several computational models have been developed to predict miRNA targets, albeit with limited performance and utility.
This study presents PRIMITI, a novel predictive model that identifies miRNA and target mRNA interactions in terms of binding activities and repression activities. The PRIMITI tool incorporates novel features, including human genetic variation and physicochemical descriptors generated by iLearn, to improve the characterisation of functional target sites positioned in the 3'-untranslated regions of mRNAs. Furthermore, a negative sample selection approach is implemented to improve the reliability of the dataset.
The excellent performance of PRIMITI was demonstrated through cross-validation and independent blind test sets. The model achieved generalisable predictive performances, with Matthew's correlation coefficients of 0.81 for identifying miRNA-target binding sites and 0.77 for predicting miRNA repression activity. The model was also validated using an external microarray dataset and experimentally verified miRNA-mRNA interactions from miRTarbase and Tarbase databases. PRIMITI outperformed existing state-of-the-art methods and demonstrated its usability in predicting miRNA-mediated repression for initial screening. PRIMITI is publicly available through a user-friendly web server at https://biosig.lab.uq.edu.au/primiti. This resource will facilitate researchers to gain more insights into the miRNA targeting mechanisms and their implications in human diseases. Consequently, it will contribute toward our understanding of miRNA function and its prospective medical application.

Ruyi Chen, The University of Queensland

Deciphering the Black Box: Mastering the MSA Transformer for Phylogenetic Tree Reconstruction

Proteins are composed of peptide bonds that link together amino acids in a sequence. By comparing protein sequences of different organisms, we can hypothesise their evolutionary relationships and shared ancestry, thereby shedding light on the functional significance and evolutionary pressures acting on those proteins. Classical methods of inferring phylogenetic relationships employ mathematical models, such as Maximum Likelihood and Bayesian inference coupled with continuous-time Markovian evolutionary models. Protein Language Models (PLMs) offer an alternative pathway to recover evolutionary relationships. Much like how natural language processing perceives sentences as chains of words, a protein sequence can be envisioned as a "sentence," with amino acids analogous to words. However, the "black box" attributes of neural networks can shroud the rationale behind their conclusions, complicating the use of PLMs in phylogenetic tree reconstruction. To this end, we illustrate how a PLM framed around a multiple-sequence alignment (MSA), the MSA transformer, encodes phylogeny despite of not being explicitly trained to recognise such, and provide a guide for phylogenetic tree reconstruction. Equipped with insights learned from the MSA transformer, we then reconstructed a phylogenetic tree for RNA virus RNA-dependent RNA polymerase

(RdRp) domain, demonstrating how both new and previously known evolutionary relationships are available from a “non-classical” approach with different computational requirements. It is anticipated that PLMs will complement classical phylogenetic approaches to accurately piece together the evolutionary history of protein families.

Yunzhuo Zhou, The University of Queensland

DDMut-PPI: predicting effects of mutations on protein-protein interactions using deep learning

In molecular biology, understanding how mutations affect protein-protein interactions (PPIs) is vital because of their impacts on diseases and implications for therapeutic design. While several prediction tools have been advanced, a prevalent issue has been their compromise between speed and accuracy. In response to this challenge, we present DDMut-PPI, a deep learning model that quickly and accurately predicts how missense mutations change protein-protein binding free energy. Building upon the robust siamese network architecture with graph-based signatures from our prior work, DDMut, the DDMut-PPI model was further augmented with a graph convolutional network (GCN) operated on the protein interaction interface. Residue-specific embeddings from the cutting-edge ESM-2 protein language model were assigned to the graph as node features, while diverse molecular interactions were used as edge features. By merging evolutionary context with spatial information, this sophisticated framework allowed DDMut-PPI to detect both proximate and distal effects of mutations, yielding a robust Pearson's correlation of up to 0.70 (RMSE: 1.46 kcal/mol) in our evaluations, outperforming most existing methods. Importantly, the model demonstrated consistent performance across mutations, irrespective of their propensity to strengthen or weaken PPIs. We believe that DDMut-PPI offers a significant advancement in the field and will serve as a valuable tool for researchers probing the complexities of protein interactions.

Ebony Watson, The University of Queensland

Overcoming Heterogeneity for Improved CNN-Based Classification of Replicative Senescence

Senescence is a cellular stress response believed to underpin biological ageing and a variety of associated diseases through its pro-inflammatory characteristics. Despite its clinical significance, identifying and characterising senescent cells within complex single-cell data has been challenging due to the diverse and dynamic nature of the senescence phenotype. Recent studies have found some success in classifying senescent cells based on morphological features through application of Convolutional Neural Networks (CNN) to high-throughput cellular microscopy data. However, these studies have used a simplified experimental model of replicative senescence, which doesn't consider the extensive temporal and intra-population heterogeneity known to characterise this cell state, and as such, limits the reliability and utility of the classifiers produced.

Here, we have improved upon these prior models to produce a highly-accurate CNN classifier for replicative senescence in Mesenchymal Stem Cells. We achieve this by imaging cells at an additional intermediate timepoint during senescence and automating per-cell labelling of senescence according to its SA-β-gal staining, rather than applying population-wide labelling based solely on timepoint. This expands our dataset for characterising the replicative senescence phenotype to three timepoints as compared to the previous single timepoint approach. Through various model comparisons, we show that accounting for temporal and intra-population heterogeneity in this way improves model accuracy by up to 63%. Moreover, models trained on timepoint-based labels of senescence produce misleading results, with recall dropping by over 25% when evaluated against the same data labelled according to SA-β-gal staining instead. Consequently, our findings highlight and provide solutions to key pitfalls in current deep learning approaches to solving biological problems.

COMBINE 2023 Symposium

Speaker abstracts

‘Omics Methodology

Carl Schmitz, Queensland University of Technology

Faster off-target risk assessment for CRISPR guide RNAs

The design of CRISPR-Cas9 guide RNAs is not trivial. In particular, evaluating the risk of off-target modifications is computationally expensive. For each 20bp candidate guide, it requires identifying among all CRISPR sites in the genome those that are at most 4 mismatches away from this guide. Large genomes have hundreds of millions of CRISPR sites, so the brute-force approach is not practical.
We previously introduced Crackling, a guide RNA design tool that indexes the CRISPR sites using 5 slices of length 4, and can generate an approximate neighbourhood for a candidate guide by requiring an exact match on at least one of the slices. We showed that this is an order of magnitude faster than other published methods. However, these approximate neighbourhoods are still large and require extensive processing to discard many ‘false’ neighbours (that are more than 4 mismatches away).
Here, we propose an enhanced approach, which relies on longer slices. This requires a larger number of slices, and also generates larger indexes. We discuss an algorithm to identify the smallest set of slices that still guarantees no false negatives, the use of memory-mapping to reduce the memory footprint, and the results of our C++ implementation. We reduce the number of comparisons 1000-fold, which significantly speeds up off-target scoring.

Caitlin Page, Peter MacCallum Cancer Centre

Damsel: an R package for end to end analysis of DamID

DamID is a powerful alternative to ChIP-seq for identifying where DNA-interacting proteins (such as transcription factors) bind to the genome. Fused to a protein of interest, Dam methylates adjacent GATC sites when the protein interacts with DNA. Once methylated, these sites are cut and sequenced, resulting in bound regions enriched relative to control samples. However, unlike ChIP-seq, DamID has very few analytical tools available, thereby limiting the ability of researchers to analyse and visualise the data in meaningful ways.
Here we introduce Damsel, the first dedicated R package enabling end to end analysis of DamID. Taking BAM files as input, Damsel counts reads mapping to each region and tests for enriched regions utilising edgeR. Damsel conducts the main steps of DamID analysis; identifying peaks and candidate gene targets, as well as performing gene ontology testing with added bias correction (the number of GATC sites per gene) from goseq. Damsel also contains a variety of ggplot layers, allowing for an IGV style visualisation of stacked results (coverage, peaks, logFC etc) from within R. Damsel performed well compared to the widely used command line tool; damidseq_pipeline (Marshall & Brand, 2015), with a large overlap between the identified peaks when applied to Drosophila Scalloped data. Overall, Damsel presents a comprehensive approach for researchers wishing to conduct an end to end analysis for DamID in one centralised place, with exploratory and visual capabilities standard to other high throughput data types.

Roxana Zahedi Nasab, The University of New South Wales

Improving Spatial Domain Clustering and Trajectory Inference with Dynamic Biologically-informed Graph Structure and Graph Convolutional Neural Network

Spatially Resolved Transcriptomics (SRT) has gained substantial interest in research due to its capacity to uncover spatial information from captured locations named "spots" at a single-cell resolution. The analysis of SRT data requires an unsupervised clustering approach to identify spatially coherent regions, known as spatial domain clustering, which is a fundamental aspect of SRT data interpretation and all the subsequent analyses. Leveraging the accessibility of spatial coordinates and employing similarity metrics to measure the distances between the spots, recent studies have increasingly explored the modelling of SRT data as a graph structure with gene expression as features. However, most studies have predominantly relied on using the gene expression matrix alone or employing a fixed graph structure with predefined neighbouring relationships to construct graph neighbourhoods, which lacks inherent biological information as different cell types may have varying degrees of interactions and dependencies. Here, we propose Dynamic Biologically-informed Graph Convolutional Neural Network for Spatial Transcriptomics analysis (DynBiST). DynBiST employs an unsupervised learning technique, incorporating the results of clustering during each learning iteration as a form of intrinsic biological information, eliminating the requirement for a predefined ground truth. By dynamically weighting each spot based on its clustering score, we iteratively refine the graph structure, resulting in a more biologically-informed graph representation. Subsequently, a graph convolutional neural network tailored with an innovative loss function is employed to extract embeddings (compact numerical representations capturing intrinsic properties) from the graph. We assessed the performance of DynBiST on the LIBD human dorsolateral prefrontal cortex and mouse olfactory bulb datasets. Our findings revealed that DynBiST significantly improve the clustering performance of spatially coherent regions compared to SpaceFlow, conST, STAGATE, and GraphST by achieving at least 10% higher clustering accuracy in both the datasets. Furthermore, the embeddings obtained by DynBiST improved downstream analyses, including trajectory inference and cell-type deconvolution.

COMBINE 2023 Symposium

Posters

1	**Natsuki Sasaki**	Naturally selected human cardiotropic adeno-associated virus to overcome translational barriers for the use in cardiac gene therapy
2	**Prakrithi Pavithra**	Unravelling LncRNA Diversity at a Single Cell Resolution and in a Spatial Context of the tumor Across Different Cancer Types
3	**Jialin Ding**	3D spatial gene expression reconstruction using single-cell RNA-sequencing data
4	**Stefanie Navaratnam**	Comparing the mutation profile of epigenetic modifier genes by evolutionary age in cancer
5	**Janice Reid**	Inhibition of BET proteins modulate transcription to prevent inflammation-driven cardiac dysfunction
6	**Di Xiao**	Refate identifies chemical compounds that target trans-regulatory networks for cellular conversion
7	**Xiaoqi Liang**	Benchmarking Spatial Transcriptomics Simulations: A comprehensive Evaluation Framework
8	**Sirui Weng**	Disentangling PSMA heterogeneity in advanced prostate cancer using single-nuclei RNA sequencing
9	**Xiaoqi Liang**	Benchmarking Spatial Transcriptomics Simulations: A comprehensive Evaluation Framework
10	**Lachlan Baer**	tadar: a Bioconductor package for Transcriptome Analysis of Differential Allelic Representation
11	**Yangyi Zhang**	Deciphering cell-cell communication patterns between the tumour and the tumour microenvironment in advanced prostate cancer
12	**Samuel C. Lee**	Modelling nanoparticle cell interactions through the protein corona
13	**Changqing Wang**	Igniting full-length isoform analysis of single-cell RNA-seq data with FLAMES
14	**Ashley L. Weir**	IdentifiHR: A gene expression logistic regression classifier to identify homologous recombination deficient ovarian cancers
15	**Guan Gui**	Incorporating sub-cellular, cellular, and super-cellular level features for imaging-based spatial genomics
16	**Anal Kanti Roy**	Contiquity: An efficient approach to accurately ensemble large set of contigs
17	**Thanushi Peiris**	Pipeline to segment and assess the morphology and patterning of kidney organoids using the Segment Anything Model

18	**Jenna Supper**	Computational Discovery of Critical Mineral Binding Proteins for Mining Applications
19	**Reza Ghamsari**	Comparative Analysis of Single-nucleus and Single-Cell RNA sequencing data: Insights and Trade-offs
20	**Aaron Kovacs**	MTR3D-AF2: Expanding the Coverage of Spatially Derived Missense Tolerance Scores Across the Human Proteome Using AlphaFold2
21	**Ha M. Tran**	Seminal fluid histocompatibility antigens contribute to T cell priming after mating in mice
22	**George Bouras**	Pharokka and Megapharokka: enabling automatic, consistent, scalable and sensitive annotation of the bacteriophage genomic universe
23	**Max Woollard**	Sliced Inverse Regression for Single Cell and Spatial Gene Expression Data with Slice Weighting
24	**Namuhan**	Benchmarking performance of scRNA-seq cell type annotation methods under imbalanced cell type proportions
25	**Sanjana Tule**	Optimal Phylogenetic Reconstruction of insertion and deletion events
26	**Farhan Ameen**	Context is important! Identifying context aware spatial relationships with Kontextual
27	**Vladimir Morozov**	Accelerating protein molecular dynamics using machine learning
28	**Rossen Zhao**	Rapid and sensitive read-based profiling of viruses with conserved sequence tags
29	**Kira Villiers**	genomicSimulation: a fast, flexible tool for stochastic simulation of breeding programs
30	**Franco Caramia**	Establishing the Link between X-Chromosome Aberrations and TP53 Status, with Breast Cancer Patient Outcomes
31	**Ashlee Thomson**	Enhancing Precision and Eliminating Reference Bias: Building a B-cell Acute Lymphoblastic Leukaemia Specific Pan-Genome Graph for Improved Genomic Alignment
32	**Eileen Ho**	Unlocking the potential of individual genetic heterogeneity responses to hypoxia for novel cardiovascular drug discovery
33	**Hsiao-Chi Liao**	Removing unwanted variation from antibody-derived tag counts with RUV-III-NB
34	**Lijia Yu**	Ensemble deep learning of embeddings for clustering multimodal single-cell omics data
35	**Daniel Kim**	Knowledge-guided single-cell clustering and cell type annotation
36	**Chuhan Wang**	Benchmarking translational potential of spatial transcriptomics imputation from histology images
37	**Chensong Chen**	Genomic prediction for sugarcane diseases including hybrid Bayesian-machine learning approaches

38	**Amanda Bataycan**	A Scoring Method to Assess the Effect of Pathogenic Single Nucleotide Variants on Protein-Coding Genes in Patients with Leukemia
39	**Elijah S Willie**	An autoencoder based framework for predicting clinical outcomes with high parameter cytometry data
40	**Yunfan Fu**	Integrating deep mutational scanning and low-throughput mutagenesis data to predict the impact of amino acid variants
41	**Zixiong Zhuang**	First study to discover and characterise population-level structural variants (SVs) in wild populations of Eucalyptus viminalis
42	**Christian Degnbol Madsen**	The Topological Properties of the Protein Universe
43	**Lijun Xu**	Multi-omic analyses reveal profiles of cancers with methylated homologous recombination genes
44	**Lachlan McKinnie**	Using a custom Python script to predict metabolites in red algae using KEGG functional annotations
45	**Brett Liddell**	Exploring the relevance of RAD51C methylation in stomach adenocarcinoma using the CCLE
46	**Sophie Leech**	Women with pre-existing type 2 diabetes mellitus have altered gut microbiome composition in pregnancy
47	**Lei Qin**	Harnessing Nearest-Neighbour Graphs to Better Understand Single-Cell Topologies and Clustering
48	**Ziming Chen**	Nanopore Library Preparation Kit Biases and Microbiome Analysis
49	**Sagrika Chugh**	Data generative model for simulating single-cell ATAC-seq data
50	**Carissa Chen**	Evaluating spatially variable gene detection methods for spatial transcriptomics data
51	**Kian Soon Hoon**	minSNPs: from derivation of resolution-optimised SNP sets to analysis of Nanopore whole genome sequence data
52	**Jinjin Chen**	Adapting natural language processing method TF-IDF for Marker Selection and Gene-Set Scoring of Single-Cell RNA-Seq Data
53	**Zoe Hunter**	Utilising transcriptomic data to uncover drug mechanisms of action
54	**Sarah Shah**	Massive genome reduction predates the divergence of Symbiodiniaceae dinoflagellates
55	**Halimat Chisom Atanda**	Long read sequencing to identify genomic and epigenomic variations between MCF7 sub cell lines
56	**Harvey Santos**	Genomic Insights into Tritrichomonas foetus and its Role in Vaccine Development: A Pore-C Approach
57	**Adam Serghini**	Prediction of pathogenic missense mutations in RET using machine learning

58	**Shweta Mall**	Whole Exome Sequence Analysis Identifies Rare Variants Associated with Milk Production Traits of Bos Indicus Cattle Breeds of Indian Sub-continent
59	**Brendan Jeon**	Genotyping structural variants with long reads, at population scale using deep leaning
60	**Weixiong He**	Polygenic risk scores predict intraocular pressure and vertical cup/disc ratio and improve the risk prediction of open-angle glaucoma
61	**Bhavya Papudeshi**	Spae, a bioinformatic workflow to rapidly detect phage therapy candidates among isolated phages
62	**Dillon Wong**	Automated Workflows for the Pre-processing and Analyses of Spatially Resolved Transcriptomics Data
63	**Andy Tran**	Lipidomic signatures of resilience to coronary artery disease learnt from chimpanzees
64	**Xiaotong Gu**	EFG-CS: Predicting Chemical Shifts from Amino Acid Sequences Using Graph Neural Networks and Transfer Modelling
65	**Natalie Charitakis**	Disparities in spatially variable gene calling highlight the need for benchmarking spatial transcriptomics methods
66	**Giulia Iacono**	Multi omics profiling identifies host-microbe interactions relevant for the stability of the transplanted lungs
67	**Po Jui Shih**	Efficient real-time selective genome sequencing on resource-constrained devices
68	**Susanna Grigson**	Phynteny: Synteny-based annotation of bacteriophage genes
69	**Hiruna Samarakoon**	Squigualiser - An interactive nanopore signal visualisation toolkit
70	**Kisaru Liyanage**	minimap2-fpga: Integrating hardware-accelerated chaining for efficient end-to-end long-read sequence mapping
71	**Sam Godwin**	Genetic Determinants of T2D and CVD: Evidence from a GWAS in Aboriginal South Australians
72	**Sanghyun Kim**	Establishing an evaluation framework for microbiome analytics
73	**Sidra Batool**	Unveiling the Secrets of QacA: Journey into the Molecular Dynamics of Multidrug Resistance in Staphylococcus aureus
74	**Muneeza Maqsood**	Casting Light on RNA-seq Splicing Event Authenticity: Leveraging Shannon Diversity Index for Discrimination
75	**Shreya Rajesh Rao**	A review of cell type annotation algorithms for single-cell technologies
76	**Jacob Meyjes-Brown**	Exploring the genetic basis of Metabolic Syndrome through whole-genome sequencing in the Norfolk Island population isolate
77	**Sara Zufan**	Towards genomics-enhanced outbreak surveillance in public health

78	**Roshan Jalaldeen**	Developing the ALLSorts gene expression classifier for single cell analysis of B cell Acute Lymphoblastic Leukemia
79	**Bethany Cross**	Characterisation of Mobile Genetic Elements Driving the Global Dissemination of a Critical Antibiotic Resistance Gene
80	**Peter Thomas**	Uncovering Genetic Variation Underlying Divergent Responses to Exercise Training in Selectively Bred Rats
81	**Ashley Lockhart**	Tank-side genomics for management and selection in aquaculture
82	**Jessica Hintzsche**	A Digital Twin to accelerate aquaculture genomic selection
83	**Cong Pham**	Evaluating phage genomes for variants dominance in IBD patient's metagenomics samples
84	**Adeline Trieu**	Essential gene regulatory network generation for early kidney development
85	**Sandeep Santhosh Kumar**	An annotation-guided deep neural-network method for factor analysis of single-cell RNA-seq data
86	**Quoc Anh Nguyen**	Cell Segmentation algorithms in Spatial Transcriptomics: An analysis and comparison between different Deep Learning models
87	**Debbie Chong**	Multi-omic analysis to uncover molecular mechanisms underlying absence epilepsy development in the GAERS model
88	**Malvika Kharbanda**	Single Cell Scoring of Molecular Phenotypes
89	**Oliver Cheng**	Flexiplex: a noise-tolerant sequence searching and demultiplexing tool for single cell data
90	**Claire Cheng**	Extend Colorectal Cancer Liver Metastasis Clonal investigation to Isoform with Long-Read RNA Sequencing
91	**Anthony Bengochea**	Filling out P450 phylogenies using transcriptomic sequences from underexplored animal phyla
92	**James Bradley**	Cancer Data Lakehouse for Precision Medicine
93	**Cui Tu**	Spatial profiling of paediatric rhabdomyosarcomas for understanding the tumour microenvironment
94	**Piyumal Demotte**	Robust Phylogenetic Dating with IQ-TREE and MCMCTree
95	**Aayushi Notra**	Single Cell analysis of an in vitro timecourse mode to study the origin of Neuroblastoma
96	**Gemma Laird**	Targeted Metagenomic Approach to Decipher Taxonomy and Metabolic Capability of Phylum Bipolaricaulota
97	**Natsuki Sasaki**	Naturally selected human cardiotropic adeno-associated virus to overcome translational barriers for the use in cardiac gene therapy
98	**Jia Yi Hee**	Peripheral blood gene expression associated with progression to type-1 diabetes mellitus in islet autoantibody seropositive children

99	**Aditya Sethi**	Developing long-read sequencing methods to map co-transcriptional m6A modification & pre-mRNA splicing dynamics at single molecule resolution
100	**Michael Nakai**	Isopod: Detecting differential isoform usage between cell types from long-read single cell data
101	**Solal Chauquet**	Transcriptomic changes as a marker of liver transplant viability during normothermic perfusion
102	**Isabella Burdon**	Setting the standard: first positive control for the sinonasal microbiome tovalidate the metagenomic workflow for taxonomic profiling
103	**Damien Cleary**	An Atlas of C1 Pathways for X

COMBINE Symposium and ABACBS Conference 2023

Sponsors

Platinum

MGI

Gold

CSL

Silver

Australian BioCommons

dug

NCI AUSTRALIA

QIMR Berghofer Medical Research Institute

QUT

The University Of Queensland AUSTRALIA

Bronze

10x GENOMICS

PacBio

Distributed by millennium science

Pawsey

nanoString

QCIF BIOINFORMATICS

MICROBA

Student Travel Fund Sponsor

SAiGENCI SOUTH AUSTRALIAN IMMUNOGENOMICS CANCER INSTITUTE

The University Of Queensland AUSTRALIA

COMBINE 2023 Symposium

COVID safety statement

The latest Queensland government regulations do not require masks to be worn indoors. However we strongly recommend you wear a mask at the COMBINE 2023 Symposium, in particular when social distancing can not be maintained. Surgical masks and hand sanitiser will be available throughout the conference venue.

Please do not attend the conference if you are unwell.