



Programme Booklet

Welcome

Welcome to the 2018 COMBINE Symposium. This annual event is an opportunity for students and early career researchers to present their work to peers in a relaxed and supportive environment.

Regards,

The COMBINE Team

Symposium Committee Members

Jiaping (Lucy) Liu (Co-chair)

Youwen Qin (Co-chair)

Luke Zappia (COMBINE president)

Dr. Melissa Davis (ABACBS Conference Organising Committee liaison)

Emma Gail

Feng Yan

Mailie Gall

Jiayuan Huang

Joanne Huynh

Holly Whitfield

Lochlan Fennell

Ning Liu

Graphic design team: Emma Gail, Mailie Gall & Jiaping (Lucy) Liu

Careers panel organisation: Mailie Gall & Jiaping (Lucy) Liu

Session chairs: Emma Gail, Holly Whitfield & Mailie Gall

Media and communications: Luke Zappia & Ning Liu

Registration desk: Youwin (Owen) Qin, Feng (Alex) Yan, Holly Whitfield & Mailie Gall

Photography: David Ma.

Social night organisation: Holly Whitfield, Jiaping (Lucy) Liu, Joanne Huynh

Sponsorship: Holly Whitfield, Jiaping (Lucy) Liu

Sponsors



MASSIVE

isCB Student
WACOUNCIL
INTERNATIONAL SOCIETY FOR COMPUTATIONAL BIOLOGY



Walter+Eliza Hall
Institute of Medical Research

DISCOVERIES FOR HUMANITY

About COMBINE

COMBINE is a student-run Australian organisation for researchers in computational biology, bioinformatics, and related fields. COMBINE is the official International Society for Computational Biology Regional Student Group for Australia and a subcommittee of The Australian Bioinformatics and Computational Biology Society (ABACBS). We aim to bring together students and early-career researchers from the computational and life sciences for networking, collaboration, and professional development.

Australia has many research institutes, each with their own cohorts of students. Aside from conferences, there are few opportunities that bring these students together, allowing them to discover the different kinds of research going on at other institutes. COMBINE aims to bridge this institutional divide by organising seminars, workshops and social events. Together, these events allow students to connect with each other and build a network in a casual environment.

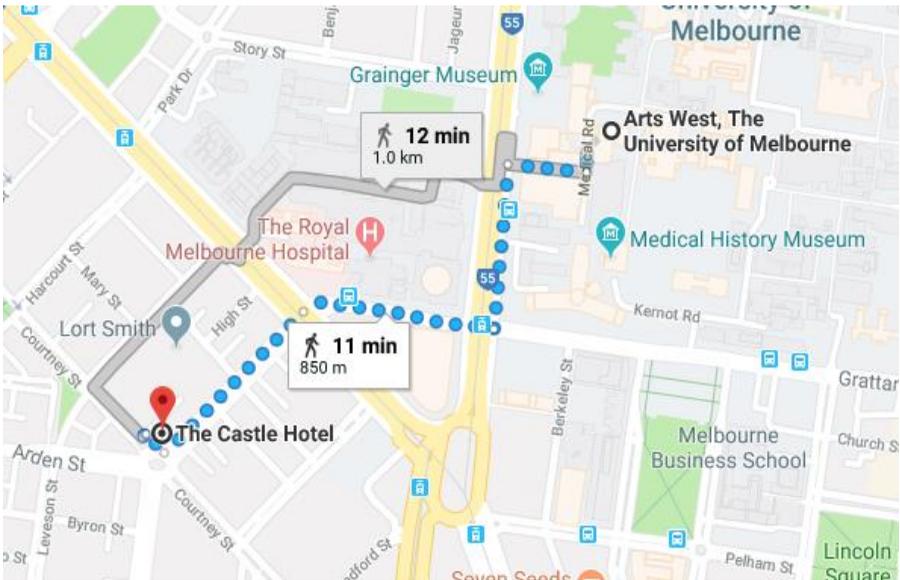
Symposium Programme

8:00	Registration and poster hanging
8:50	Symposium Welcoming Address
9:00	Matthew Z DeMaere <i>bin3C : Hi-C mediated retrieval of metagenome-assembled genomes (MAGs)</i>
9:20	Holly Whitfield <i>MicroRNA-Mediated Regulatory Networks within Breast Cancer Progression</i>
9:40	Brendan Robert E. Ansell <i>Machine learning and protein structure prediction to support annotation of pathogen genomes</i>
10:00	Jamie-Lee Thompson <i>Transcriptome-wide RNA sequencing analysis of immobilisation-induced muscle atrophy</i>
10:20	Morning tea
10:40	Peter Georgeson <i>Identifying and characterising high resolution mutational signatures from DNA mismatch repair deficient tumours</i>
11:00	Emma Gail <i>Systematic mapping of molecular interactions within the epigenetic modifier complex PRC2 provides a mechanistic framework for its functional diversity</i>
11:20	Jake Bradford <i>A Performance Review of Computational Tools for CRISPR-Cas9 Guide Design</i>

11:40	<p>Katarina Stuart <i>Evolution in invasive populations: using genomics to reveal drivers of invasion success in the Australian European starling (<i>Sturnus vulgaris</i>) introduction across Australia.</i></p>
12:00	Lightning talks - 1min talks about posters
12:20	Lunch
12:50	Poster session
13:40	Group Photo
13:55	<p>Dharmesh Bhuvu <i>Evaluation and characterisation of differential co-expression analysis methods</i></p>
14:15	<p>Vindhya Vasini Shatdarsanam <i>Identifying epistasis underlying Age-related Macular Degeneration (AMD) to understand the genetic architecture of the disease</i></p>
14:35	<p>Andrian Yang <i>Scavenger: A pipeline for recovery of unaligned reads utilising similarity with aligned reads</i></p>
14:55	<p>Mengbo Li <i>Using Omics Data to Guide Classification in Neuroimaging Studies of Brain Diseases</i></p>
15:15	Afternoon tea (with career panelists)
15:50	<p>Careers Panel Roxane Legaie, Chris Saunders, Ben Goudey and Ann-Marie Patch</p>
17:00	Address from COMBINE president Luke Zappia

17:20	Awards - Best posters, best talks and best lighting talk
18:00	ABACBS awards
19:15	Social night at The Castle Hotel

At **7:15 pm** we will be changing venue to The Castle our evening social event. Groups will be led over by the symposium organisers, but walking directions are also provided:



Abstracts

Oral Presentations

Matthew Z DeMaere

bin3C : Hi-C mediated retrieval of metagenome-assembled genomes (MAGs)

Of the Earth's total microbial diversity, metagenomics provides a means of studying both organisms which are readily cultured and the significant proportion which is not accessible through culture-based methodologies. An emerging refinement within this field is the systematic extraction of Metagenome-Assembled Genomes (MAGs), representing the resolution of individual genomes from whole metagenomes.

Many leading approaches to effective MAG retrieval rely on the production of multi-sample time-series data, the cost of which may be burdensome. Additionally, multi-sample studies can pose a logistical barrier in some settings, such as with clinical isolates. Eliminating the need for time-series data, the high-throughput sequencing technique called Hi-C is a promising alternative approach.

The Hi-C technique produces read-pairs which capture in-vivo DNA-DNA proximity interactions. Within a metagenome, the physical structure of the community modulates the signal derived from these interactions and a hierarchy of interaction rates exists (Intra-chromosomal > Inter-chromosomal > Inter-cellular).

We describe bin3C, an unsupervised method that exploits the hierarchical structure of Hi-C interaction rates to resolve MAGs from a single time-point. Using ground truth validation of a simulated microbiome the method is shown to possess high precision and good recall. As the only Hi-C based alternative to MAG retrieval, bin3C is compared directly against the recently announced proprietary service ProxiMeta. bin3C performed favourably against ProxiMeta when analysing a real human faecal microbiome, retrieving 55 vs 35 nearly-complete genomes and with a lower median contamination rate of 0.8% vs 3.5%.

Holly Whitfield

MicroRNA-Mediated Regulatory Networks within Breast Cancer Progression

The leading cause of death for breast cancer patients is cancer cell metastasis, mediated in part by Epithelial-to-mesenchymal transition (EMT), a regulatory program controlling cell phenotype. There is growing evidence that microRNAs (miRNAs), particularly the miR-200 family, play a central role in EMT regulatory networks that underlie breast cancer progression. The relationship between miRNAs and their targets cannot always be directly identified. Experimental approaches often require the isolation and independence of molecular pathways, disregarding much of their broader biological context.

MicroRNAs exert control over cellular phenotypes by the coordinated effects of multiple miRNA, which includes the additive effects of multiple miRNA co-targeting individual mRNA, as well as a single miRNA targeting multiple mRNAs. Transcripts with numerous binding sites can ‘sponge up’ miRNA, decreasing the availability for other transcripts. It is in this way that transcripts compete for miRNA and indirectly regulate one another, and is referred to as the competing endogenous RNA (ceRNA) hypothesis.

CircularRNA (CircRNA) are circularised transcripts that are protected from degradation, and often contain many microRNA binding sites. The unexpected dependencies between competing genes may be explained by Quaking-mediated circularisation of transcripts. Here, we have developed a workflow that integrates TargetsScan and DIANA-microT predictions, a cell line model of EMT, as well as TCGA breast cancer tumour samples, to identified relationships that best describe the stoichiometry between circularised transcripts and EMT-associated microRNAs. In particular, we propose a mechanism through which EMT is induced by the dysregulation of the canonical epithelial phenotypical regulatory network.

Brendan Robert E. Ansell

Machine learning and protein structure prediction to support annotation of pathogen genomes

Large-scale computational prediction of protein structures represents a cost-effective alternative to empirical structure determination with promise for under-studied human pathogens. Conventional sequence-based tools are insufficient to annotate the genomes of such divergent biological systems. Conversely, protein structure tolerates substantial variation in primary amino acid sequence, and is thus a highly sensitive indicator of biochemical function. Structural proteomics is poised to become a standard part of pathogen genomics research, however informatic methods are now required to assign confidence in large volumes of predicted structures. We used the I-TASSER suite to predict structures for ~5000 proteins encoded in *Giardia duodenalis* and identified their closest empirically determined structural homologues in the Protein-DataBank. Models were assigned to high or lower-confidence categories depending on the presence of matching PFAM domains in query and reference peptides. Metrics output from I-TASSER and derived metrics were assessed for their ability to predict the high confidence category individually, and in combination through development of a random forest classifier. Predicted structures and confidence predictions are freely available at predictein.org/giardia_duodenalis.

We identified 1115 high confidence models including 221 hypothetical proteins. Amino acid identity between query and reference peptides was the greatest individual predictor of high confidence status, however the random forest classifier out-performed any metric in isolation (AUC = 0.972), and identified a subset of 142 high confidence-like models (i.e., false-positives). High confidence models exhibited higher transcriptional abundance, and the classifier generalized across species, indicating the broad utility of this approach for automatically stratifying predicted structures.

Jamie-Lee Thompson

Transcriptome-wide RNA sequencing analysis of immobilisation-induced muscle atrophy

Disuse atrophy is a secondary complication that often exacerbates the aetiology of injury and chronic disease. Identifying changes in mechanisms that control muscle mass is necessary to characterize atrophy and maintain a healthy functional capacity. This study aimed to use RNA sequencing as a high resolution, untargeted approach to study gene expression in human skeletal muscle following short-term limb immobilisation. Skeletal muscle biopsies were collected from the m. vastus lateralis before and after 3 days and 14 days of immobilisation (63 samples total). Skeletal muscle RNA was isolated and analysed using Illumina RNA sequencing. Initial data analysis compared Salmon, a quasi-alignment tool and STAR, a traditional alignment tool to identify the most appropriate alignment method for this dataset. Mapping of all samples showed Salmon performed much faster than STAR and had a higher percentage of reads mapping back to the transcriptome (77.3% to 67.4%), while quality control metrics were comparable. Gene set enrichment analysis of RNA sequencing data identified significant changes in mitochondrial processes and carbohydrate metabolism, which became more pronounced throughout immobilisation. Furthermore, changes in chromatin remodelling processes appeared to be transient changes, as they were only observed at 3 days. Conversely, changes in protein localisation, ribosome biogenesis, translation, nucleotide metabolism and oxidoreductase activity were only identified after 14 days. This research is the first extensive transcriptomic study of short-term disuse atrophy in human skeletal muscle. These data support mitochondrial dysfunction and reduced protein synthesis as key events during muscle disuse atrophy.

Peter Georgeson

Identifying and characterising high resolution mutational signatures from DNA mismatch repair deficient tumours

Microsatellite instability (MSI) is a hypermutated phenotype commonly observed in several cancers and is associated with loss of functionality in the mismatch repair (MMR) mechanism. Identifying tumour MSI has important clinical implications, affecting therapeutic choices, prognosis, and familial cancer risk appraisal.

Traditionally, methods of MSI detection have been based on measuring changes in a small panel of microsatellite tandem repeat markers, but this approach has recently been expanded to utilize next generation sequencing (NGS) data to include many more informative markers, enabling a more comprehensive examination of the MSI genomic signature.

MSI does not affect the genome uniformly; it is shaped both by the evolution of the tumour, and the nature of the MMR loss. In addition, many MSI tumours exhibit both common and unique mutational signatures, but there is currently no existing method for differential analysis of cohorts.

To address these limitations, we have developed a new bioinformatics workflow to assess MSI in both whole exome sequencing (WES) and whole genome sequencing (WGS) data. By considering both the genomic loci and sequence context of mutations, MSI-diff enables the identification and comparison of MSI signatures across collections of samples. We have applied MSI-diff to colorectal and prostate cancer samples with the aim of identifying subtypes of MSI with clinical utility, including differentiation of MSI signatures between inherited and sporadic cancers. Here we present the overall algorithm underpinning our workflow and illustrate its application in the context of two cancer types.

MSI-diff is publicly available at <https://github.com/supernifty/msi-diff>

Emma Gail

Systematic mapping of molecular interactions within the epigenetic modifier complex PRC2 provides a mechanistic framework for its functional diversity

The polycomb repressive complex 2 (PRC2) is a histone methyltransferase complex that tri-methylates histone H3 at lysine 27 (H3K27me3), an epigenetic repressive mark that is essential to maintain the repressed state of thousands of genes during development. Interactions between PRC2 core subunits (EZH2, SUZ12, EED, and RBBP4) to its various accessory subunits and RNA provide it with extensive functional diversity. Mapping the domains and surfaces within PRC2 that are utilized for most of these interactions would lead to discovering how the function of PRC2 is regulated by these interactions and how these interactions are dysregulated in disease. To this end, we developed workflows to automatically detect, filter and aggregate binding sites from two types of cross-linking mass spectrometry (XL-MS) experiments to study protein-protein and protein-RNA interactions. Cross-linking with mass spectrometry (XL-MS) used to map protein-protein and protein-RNA interactions between PRC2 to its co-factors and ligands. The most common protein-protein interactions within subunits of the core PRC2 complex varied slightly between the presence and absence of various protein cofactors. We identified domains and surfaces in PRC2 that bind different proteins and RNA and provide a mechanistic framework to explain why certain factors are mutually exclusive for PRC2 binding while others can co-occupy the same complex. Given that >1000 amino acids within PRC2 subunits identified with disease-associated mutations – mostly in cancer – this workflow can help determine how protein-protein and protein-RNA interactions are altered during disease and what are the probable functional consequences of these mutations.

Jake Bradford

A Performance Review of Computational Tools for CRISPR-Cas9 Guide Design

The difficult computational task of designing safe and efficient CRISPR-Cas9 guides for genome editing has challenged bioinformaticians. Due to this, a myriad of software tools exist, however, no detailed analysis of their performance and output has been reported. Here, we review ten existing tools and analyse how system resources are utilised, how run time responds to an increasing input size and we perform a detailed analysis of the output generated by each tool. To achieve this, we implemented a system-auditing method and tested each tool on up to four datasets of increasing size (from 500 thousand to 61.4 million base pairs, derived from GRCm38/mm10 chromosome 19). We found that six tools are not practical to analyse a whole genome: four tools could not complete the analysis of one chromosome within a 72 hour time limit, and two tools triggered the operating system's out-of-memory killer. Most tools reported guides that had perfect matches elsewhere in the dataset being processed, relying on the user to determine the appropriateness of a guide. On the other hand, one tool (mm10db) uses a strict rule-set to filter out inappropriate candidate guides. Combining this filtering process with some of the features of the fastest tools (CasFinder and CRISPR-ERA) provides a direction for the development of a computational design method that will produce safe and efficient guides at a rate practical for the reliable analysis of entire genomes.

Katarina Stuart

Evolution in invasive populations: using genomics to reveal drivers of invasion success in the Australian European starling (Sturnus vulgaris) introduction across Australia.

Invasive species are a global concern due to their negative impacts on the economy and local ecosystems. However, well documented invasions provide a useful system in which to pose biologically interesting questions regarding short time scale evolution of species. The answers to these questions may then further our knowledge of evolutionary mechanism, as well as inform specific management strategies for the invasive population. The global pest European starling (*Sturnus vulgaris*) was introduced into Australia's south-eastern states in the 1860's and have since greatly expanded their range. Previous research on multiple introduced starling populations has demonstrated that their morphology has undergone subtle shifts following colonisation. Research is currently ongoing to characterise the global genetic differences between the contemporary native population in the UK, and two of the major continental invasions, Australia and North America. Our study sought to investigate the intracontinental whole genome variation across Australia's starling population. Whole genome resequencing data was obtained for 14 starlings from key geographic sites across the Australian invasion (e.g. introduction site, invasion edge). Analysis of the whole genome variation between these subpopulations will reveal possible signatures of selection in this species. Comparing variation to the annotated reference genome will help identify possible function SNPs that may be driving the phenotypic divergence across the Australian population.

Dharmesh Bhuva

Evaluation and characterisation of differential co-expression analysis methods

Elucidation of regulatory networks is a key aim in systems biology, including identification of regulatory mechanisms specific to a given biological context. This has motivated the move from co-expression to differential co-expression analysis and numerous methods have been developed subsequently to address this problem. Evaluation of methods and interpretation of resulting networks has been hindered by the lack of known context-specific regulatory interactions.

We previously developed a simulator based on dynamical systems modelling capable of simulating differential co-expression patterns. Recently we integrated the simulator into an evaluation framework to benchmark and characterise the performance of inference methods. Defining three different levels of “true” networks for each simulation, we showed that accurate inference of causation was difficult for all methods, compared to inference of associations. We show that a z-score based method had the best general performance. Further, analysis of simulation parameters revealed five properties that explained the performance of methods.

Analysis of inferred networks showed how hub nodes were more likely to be differentially regulated targets than transcription factors and proposed an interpretation of the inferred differential network based on this observation which helps us reconstruct a putative causative network. Application to a breast cancer dataset revealed differential regulation of immune processes dependent on estrogen receptor status.

The potential of differential co-expression analysis remains largely unexplored due to difficulties in interpreting results, and we have attempted to address some of the limiting factors. Applications of methods are not limited to co-expression and may be applied to associations in general.

Vindhya Vasini Shatdarsanam

Identifying epistasis underlying Age-related Macular Degeneration (AMD) to understand the genetic architecture of the disease

Background: Gene-gene (GxG) interactions help in estimating the missing heritability component in genome-wide association studies. AMD is a perfect platform to study the effect of statistical interactions as approximately 50% of genetic variance has already been explained by genetic associations. Methods: The main resource for this study arises from the 40,000 AMD cases and controls, distributed across 26 cohorts and genotyped by the International AMD gene Consortium(IAMDGC). Seven of these cohorts were available locally on our servers and the genomic regions of the 34 loci previously identified by the IAMDGC as associated with AMD were considered to assess GxG interactions. To identify this, three popular epistasis tools namely PLINK, BOOST and Antepiseeker were used. Consistency of interactions across the 7 AMD cohorts was used as a criteria to investigate the efficiency of these tools.

Results: Using PLINK, we identified two intronic single nucleotide polymorphisms (SNPs) (rs6695321 & rs424535) from the CFH gene showing an additive effect and replicated across 6 of 7 cohorts with a combined Fisher's p-value of 1×10^{-46} . Seven other CFH-CFH interactions were observed across 5 cohorts. Using BOOST, 10 interactions were identified across 5 cohorts and 12 interactions consistent across 4 cohorts with the top interaction being between seq-rs7513157 and kgp15357111($p=1 \times 10^{-39}$). Using Antepiseeker there was only one interaction between rs7540032 and rs3750847 consistent across two cohorts. Conclusion: In this analysis on cross-Cohort consistency, PLINK showed the highest while Antepiseeker least consistency. Variability between epistatic techniques appears to be a limitation in this field.

Andrian Yang

Scavenger: A pipeline for recovery of unaligned reads utilising similarity with aligned reads

Read alignment is an important step in RNA-seq analysis as the result of alignment forms the basis for further downstream analysis. However, recent studies have shown that published alignment tools have variable mapping sensitivity and do not necessarily align reads which should have been aligned, a problem we termed as the false-negative non-alignment problem. We have developed Scavenger, a pipeline for recovering unaligned reads using a novel mechanism which utilises information from aligned reads. Scavenger performs recovery of unaligned reads by re-aligning unaligned reads against a putative location derived from aligned reads with sequence similarity against unaligned reads. We show that Scavenger can successfully recover unaligned reads in both simulated and real RNA-seq datasets. We also compared the performance of our pipeline against the Read Origin Protocol (ROP) tool for recovering unmapped reads and show that reads recovered by Scavenger have less false positive alignments compared to ROP.

Mengbo Li

Using Omics Data to Guide Classification in Neuroimaging Studies of Brain Diseases

Gene expression studies are playing an increasingly important role in understanding human brain functions and will ultimately contribute to the development of therapies and treatments to neurological diseases (Sunkin et al., 2013). The Allen Brain Atlas is a collection of data sets providing extensive gene expression, connectivity and neuroanatomical data, including spatially mapped microarray data from human brain. In this presentation, we will present an integrative approach for multi-model data from spatial prospective that utilises the imaging and omics data from publicly available brain researches. Instead of focusing on matching sample information in multi-model integration, we propose a method focusing on functional brain structure matching to enable integrative analysis of multi-modal data. We will describe the construction of various gene sets and how to incorporate these into the feature selection for classification of fMRI data. Finally, we will describe an evaluation framework to compare and contrast the effectiveness of the various types of gene sets.

Poster Presentations

Posters - Lightning Talk

1 **Melanie Smith**

A comprehensive miRSeq profile of miRNA in the human placenta across early gestation.

2 **Tingting Gong**

A critical look at somatic structural variant detection for cancer genomics

3 **Ning Liu**

A glance at the 3D genome structure of regulatory T cells with in situ HiC

4 **Tulio Campos**

A systematic evaluation of eukaryotic essential gene predictions using machine-learning algorithms trained with protein sequence-derived features

5 **Heeva Baharlou**

Autofluorescence Remover: A Novel Method to Identify and Remove Tissue Autofluorescence

6 **Qing Wang**

Comprehensive evaluation of somatic variant detection algorithms using Ion Torrent targeted deep sequencing data

7 **Rick M Tankard**

Detecting disease-causing repeat expansions in next-generation sequencing data

8 Jiaan Yu

Developing a clinically oriented workflow using whole genome data for inherited cardiac disorders

9 Anushi

Evaluation of De Novo mutation Calling Tools

10 Eddie K.K. Ip

Genetic burden in a Whole Genome Sequenced heterogeneous cohort of Congenital Heart Disease cases.

11 Nicolas Canete

High-throughput analysis of multidimensional microscopy to visualise HIV and its target cell interactions in situ

12 Ali Mahmoudi

Inference under the coalescent with recombination using a new data structure

13 Hannah Huckstep

Signalling Networks in the Analysis of Proteomic Data

14 Agus Hartoyo

Sloppy parameters in fitting a non-linear model for brain dynamics

15 Sara Ettamimi

Taxonomic characterization and functional analysis of microbial diversity in Moroccan rivers using a metagenomics approach

16 Ying Wong

The identification of genetic and epigenetic changes that contribute to T1D (Type 1 Diabetes) by ATAC-seq (Assay of Transposase Accessible Chromatin with high throughput sequencing)

Posters (cont.)

17 Anna Quagliari

Correcting unwanted variation in RNA sequencing data derived from a multi-centre study of leukemia

18 Regan Hayward

Epigenetic changes in Chlamydia-infected host cells

19 Anna S. Trigos

Genomic drivers of the fragmentation of co-expression modules regulating multicellularity in cancer

20 Eleni-Maria Michanetzi

GidB: A Mutational Hotspot for Streptomycin Resistance in Tuberculosis

21 Taiyun Kim

Impact of similarity metrics on single-cell RNA-seq data clustering

22 Brendan Robert E. Ansell

Leveraging public transcriptomics data to understand retinal health and disease

23 Kirsti Paulsen

Optimising intrinsic protein disorder prediction for short linear motif discovery

24 Alex Tokolyi

Plasmid classification and investigation through network analysis of co-occurring genes

25 Kevin Wang

RUV-Pro: Remove Unwanted Variation in prospective omics experiments

26 Yingxin Lin

scMerge: Integration of multiple single-cell transcriptomics datasets leveraging stable expression and pseudo-replication

Career Panel

Profiles

Roxane Legaie

Roxane has over 10 years experience working as a bioinformatician. Since completing a Master of Science (Informatique and Bioinformatique) in France in 2007, Roxane has worked in various institutions across Europe and Australia.



Roxane is highly experienced in building clinical bioinformatics pipelines, with her roles often requiring strong collaboration between people from research, the clinic and industry.

Roxane is currently the Lead Clinical Bioinformatician at the Peter MacCallum Cancer Centre. Her work is at the cutting edge of cancer genomics, implementing targeted next generation sequencing panels into clinical practice for the diagnosis of cancer patients. Prior to working at Peter Mac, Roxane was a Senior Clinical Bioinformatician at Monash Health.

In addition to being a highly-successful professional bioinformatician, Roxane is also a strong advocate and passionate about the translation of research into clinical practice. She sits on several professional committees, including Women in Technology, R-Ladies Melbourne and is the Professional Bioinformatics representative for ABACBS.

Ann-Marie Patch

Ann-Marie is team head of the Clinical Genomics group at QIMR Berghofer. She is an expert in cancer genomics, with her current work focusing on identifying genomic heterogeneity of cancer and its effect on patient response



to therapy. Other research interests include exploring the molecular basis of melanoma and mesothelioma. Ann-Marie is well-published across many high-impact journals including Science, Nature and PNAS with over 10,000 citations over the course of her career.

Ann-Marie has been the lead bioinformatician on several large cancer projects including the Australian Ovarian International Cancer Genome Consortium project and the Australian Mesothelioma Genomics project. Before moving to Australia, she was a post-doctoral fellow researching the genetics of inherited human diseases and was a bioinformatician within a NHS diagnostic medical genetics laboratory in the UK. Ann-Marie is active in training upcoming bioinformaticians and is part of the Bioinformatics training team for Bioplatforms Australia, delivering cancer genomics and the EMBL-EBI associated train-the-trainer courses.

Ben Goudey

Ben completed his PhD in Computer Science at the University of Melbourne in 2016. The focus of his PhD was on exploring novel computational and statistical methods to better identify associations between genetic variants and disease.

Ben currently works at IBM

Research-Australia in Melbourne and holds an honorary appointment at the Centre for Epidemiology & Biostatistics at the University of Melbourne. He works in the Biomedical Data Science team and has two main areas of interest. The first is the development of novel machine learning and statistical methods and tools that help uncover the genotype-phenotype relationship from large-scale genomic data. The second is the development of robust and scalable tools for analysing with whole genome sequencing data that can enable the shifting of genomics from a research tool to a solution for applications in industry.

Ben has published across a broad range of journals targeted at bioinformatics, computer science and health science disciplines. In addition to being an interesting and successful early career researcher, he also holds US patents, one relating to a computer method to identify interacting DNA loci, the other for the diagnosis of heart disease using novel lipid biomarkers.



Chris Saunders

Chris graduated from the University of Washington in 2007 with a Ph.D in Computational and Molecular Biotechnology. He joined Illumina's bioinformatics team in 2008 where he has been involved in a variety of primary and secondary analysis analysis efforts for Illumina's sequencing products, including development of the Strelka somatic variant caller. More recently, Chris has been leading an Illumina team focused on developing efficient, automated secondary analysis methods to accelerate the application of whole genome sequencing to the treatment of cancer and rare genetic disease. These methods include the Manta structural variant caller, released in 2015, and the Strelka2 small variant caller recently published in Nature Methods.

