AMSI

19

BIOINFO SUMMER

A SYMPOSIUM IN BIOINFORMATICS

CHARLES PERKINS CENTRE
THE UNIVERSITY OF SYDNEY
2–6 DECEMBER

POSTER ABSTRACTS

AMSI AUSTRALIAN MATHEMATICAL SCIENCES INSTITUTE
Australian Government Department of Education
THE UNIVERSITY OF SYDNEY

A PORTABLE BIOINFORMATICS PIPELINE FOR THE ANALYSIS OF LARGE-SCALE RNA-SEQ DATASETS

Hamish Mundell, The University of Sydney

In recent years, Next Generation Sequencing has led to the generation of several very large datasets. These datasets require complex processing steps to produce a meaningful output and thereby enable powerful downstream analysis. A lot of these data sources are publicly available, but due to their inherent processing complexity, they are an untapped resource. There is the potential to conduct large-scale studies which bring together multiple, related datasets from a number of studies: For example, WGS data transformed into RNA-Seq data, which can be used to later train models which distinguish Parkinson's Disease (PD) patients from healthy controls. We have developed several cutting edge bioinformatic pipelines, which seek to automate these processing steps in a computationally fast and efficient manner, while also being scalable and reusable across multiple platforms and datasets. Here we introduce one such bioinformatics pipeline, developed to enable the retrieval, processing and transcriptomic analysis of ~4000 RNA Sequencing samples from multiple different studies. We use the programs Jenkins, Cromwell and Kallisto in the development of this pipeline and demonstrate its application to the datasets of interest.

LOSS OF FUNCTIONAL GASTRIC GENES IN THE MONOTREME LINEAGE

Natasha Bradley, The University of Adelaide

Monotremes (platypus and echidna) are the oldest surviving mammalian linage, having diverged from therian mammals 190 million years ago. Monotremes have unique biology, combining elements of reptiles, birds and mammals. Interestingly, the stomach of monotremes is very small and the gastric juice of the monotreme stomach has a neutral pH of 6.2-7.4, while other mammals are acidic. Additionally, monotremes are missing all glands in the stomach, except for the Brunner's glands. This renders the monotreme stomach non-functional (in comparison to other mammalian stomachs) and appears an elongated oesophagus.

There was no molecular evidence of these observations until the publishing of the platypus genome in 2008, where various genes involved in stomach function were found to be missing or pseudogenised. These genes were involved in gastric acid secretion and protein degradation. Through searching an unpublished draft echidna genome and improved platypus genome, we have found the loss of these genes occurred not only in the platypus, but also in the echidna. This indicates that the loss of gastric genes occurred in the monotreme ancestor, giving insight into their evolution.

AN INTEGRATIVE APPROACH TO TISSUE-SPECIFIC EFFECTS OF MICRO-RNA REGULATORY NETWORKS

Tânia Marques, Faculty of Sciences / University of Lisbon

Dr Nham Tran, University of Technology Sydney

Dr Margarida Gama-Carvalho, Faculty of Sciences / University of Lisbon

miRNAs are small noncoding RNAs with role in post-transcriptional regulation of gene expression. Even though the basic mechanisms for miRNA action have been described, we are still unable to efficiently predict their impact on cellular function. When predicting targets for a miRNA, the context in which it is expressed needs to be accounted for. The aim of this work is to comprehensively understand the interaction dynamics between miRNAs and their target transcriptome across different tissues.

Paired miRNA and mRNA normalized count data corresponding to normal samples of seven tissues present in TCGA were downloaded. The samples were clustered to understand if any outliers were present. The profile of miRNA expression was assessed, and categories to include the miRNA or transcripts were created according to their level of expression and tissue specificity. The interactions of miRNAs and their targets were investigated through correlation analysis and compared across tissues. The clustering analysis revealed that, overall, the tissues cluster well together. The categories and levels of expression of miRNA do not reflect a higher number of negatively correlated transcripts (-0.5). The results of this work provide a deeper understanding of the regulatory networks governing gene expression regulation and allow further explorations of miRNA action.

DEEP ANALYSIS OF MIRNA AND ISOMIR EXPRESSION ON A SINGLE CELL LEVEL

Christopher Smith, University of Technology Sydney

Professor Gyorgy Hutvagner, University of Technology Sydney

miRNAs are small non-coding RNAs that play a role in the post-transcriptional regulation of genes. Studies have shown that miRNAs from the same precursor can vary in their exact sequence due to cellular mechanisms such as alternative drosha/dicer trimming or untemplated nucleotide additions, which may affect the miRNAs stability or target genes. Pioneering studies have already used single cell RNA-seq to highlight the complex heterogeneity of gene expression in cells or tissues previously assumed to be homogenous, and cancer research is beginning to focus on using this knowledge to develop novel treatments that can target different cell populations in tumors. Despite recent technological developments enabling single cell small RNA sequencing, very few studies have investigated how miRNAs and isomiRs are expressed on single cell level, and what role they may have in contributing to cell heterogeneity in healthy or diseased tissue remains unanswered. In this study we apply bioinformatics to published single cell small-RNA datasets to investigate miRNA and isomiR expression on a single cell level and evaluate the potential for using isomiRs as predictors of cell identity.

SINGLE CELL RNA-SEQ ANALYSIS REVEALS THE HETEROGENEITY OF THE COLONIC MESENCHYMAL CELLS IN INFLAMMATORY BOWEL DISEASE

Junwei Wang, The University of Adelaide

Dr Stephen Pederson, The University of Adelaide

Inflammatory bowel disease (IBD), with ulcerative colitis and Crohn's disease as two major forms, is characterized by chronic relapsing intestinal inflammation, and has been a worldwide healthcare problem with increasing incidence and prevalence. Its specific etiology is so far poorly understood. Mesenchymal cells of the intestinal lamina propria play important roles in immune homeostasis, and epithelial barrier maintenance. Their function was impaired in IBD through poorly defined pathways. Colonic mesenchymal cells are highly heterogeneous with overlapped marker genes, which prevented the study of their cell-type specific functions and attributes to IBD. We used a public colonic mesenchymal scRNA-seq dataset, to reveal the heterogeneity of colonic mesenchymal cells in IBD. We observed differential gene expression between stromal cells from healthy and colitis. Stromal cells with 5 subtypes from the mouse dataset, and 3 subtypes from the human data was detected. Also, 2 subtypes from mouse data, and 1 subtype from human data were detected to be the colitis-associated, with the rest shown to be healthy-associated. We also identified specific marker genes for each of these stromal subtypes. Our finding will assist further investigation of specific attributes of mesenchymal stromal subtypes, and enhancing immune monitoring of existing and new therapies in IBD.

TRANSCRIPTOMICS IN NEURODEGENERATIVE DISEASE

Kosar Hooshmand, The University of Sydney

Neurodegenerative Disease (ND) is an umbrella term for a range of conditions which primarily affect the neurons of the human brain. So far there are no effective biomarkers for monitoring NDs and their progression. A multi modal investigative approach is required for identification of specific clinically relevant biomarkers, necessary for therapeutic advancement in NDs.To identify the genetic/biological processes, which ultimately lead to the development of target therapeutics in NDs we are developing a platform which combines and analyses multi RNA-Seq data. Large downloads of multiple high-content data and statistical analysis models for these large datasets will enhance knowledge of core differentially expressed molecules and similarly perturbed pathways for their potential as biomarkers of NDs both in general and specific to different disorders, as well as appropriate therapeutic targets. Our group has dedicated computing resources, tools and workflows(Workflow Description Language (WDL) & the Common Workflow Language (CWL)) to process and analyse large and complex biological datasets (RNA sequence analysis) using machine learning/artificial intelligence and feature selection methods for identification of genetic or structural variants, expression changes and distorted alternative splicing events in relevant brain regions that can serve as potential biomarkers for the diagnosis and prognosis of various diseases including NDs.

INTROME: IDENTIFYING SPLICE-ALTERING VARIANTS AS DRIVERS OF HIGH-RISK PAEDIATRIC CANCER

Patricia Sullivan, Children's Cancer Institute

Genetic variants that affect pre-mRNA splicing can have a substantial impact on the resulting protein. Identifying and predicting a variant's impact on splicing is challenging, and current bioinformatic methodologies frequently miss these potentially medically-relevant variants.

To address this area of need, we have created and optimised Introme, a bioinformatic tool designed to identify and predict the impact of splice-altering variants. Introme uses machine learning (C5.0) to integrate predictions from multiple splice scoring tools, evaluating the likelihood of a variant to impact splicing. We applied Introme to 250 paediatric cancer patients, analysing a subset of known cancer genes in the germline and tumour WGS results.

We curated a list of 802 splice-altering variants from 150 papers to form the training and validation set for machine learning. Introme achieved the best performance (AUC: 0.96) of the tools evaluated, followed by SpliceAI (AUC: 0.93) and MMSplice (AUC: 0.81). Introme's predictions have led to the identification of 140 RNA-seq validated splice-altering variants in patients with paediatric cancer.

The application of Introme to patient sequencing data uncovers aberrations that were missed by previous analysis methods. Detecting these splice-altering variants has aided the identification of medically-relevant variants and facilitates the recommendation of personalised treatment options.

FINDING ALTERNATIVE POLYADENYLATION SIGNATURES AS CANCER BIOMARKERS

Nitika Kandhari, Monash University

Dr Paul Harrison, Monash University

Associate Professor Kaylene Simpson, The University of Melbourne

Associate Professor David Powell, Monash University

Associate Professor Traude Beilharz, Monash University

The output of cellular transcription is diversified by alternative RNA processing. For example, in addition to differential splicing, 70% of mammalian genes undergo Alternative Polyadenylation (APA). This changes the architecture of 3'-UnTranslated Regions (UTRs) and associated post-transcriptional regulatory control of mRNA fate. Short 3'UTRs are generally associated with de-differentiated proliferative cells (e.g. stem cells) whereas longer 3'-UTRs associate with more complex regulation and cellular specialisation. The literature suggests that ~91% APA genes switch to shorter mRNA isoforms in tumour cells. Our study aims to detect a signature of APA changes that are specific to triple-negative breast cancer (TNBC) that could be applied as a novel prognostic biomarker in early-stage breast cancer. Using bioinformatic analyses of 3'-focused RNA-seq approaches we studied the landscape of transcription and APA in three cancer cell lines in response to loss of PCF11, a core regulator of 3'-end formation. This shows a conservation of an expression and processing response to loss of 3'-end processing machinery. In addition to gene expression changes, we identify 3'-end "shifted" genes that are common to all 3 cell lines. We will present our current work around the idea that systematic lengthening of 3'UTR might normalise deregulated gene expression in cancer.

INVESTIGATING CHEMORESISTANCE MECHANISMS IN NEUROBLASTOMA USING A LONGITUDINAL CISPLATIN RESISTANCE MODEL

Janith Seneviratne, The University of New South Wales
Dr Anushree Balachandran, Children's Cancer Institute
Mrs Claudia Flemming, Children's Cancer Institute
Dr Marion Le Grand, Children's Cancer Institute
Professor Maria Kavallaris, Children's Cancer Institute
Professor Glenn Marshall, Children's Cancer Institute
Dr Belamy Cheung, Children's Cancer Institute
Dr Daniel Carter, Children's Cancer Institute

Neuroblastoma is the most common extracranial solid tumour in children. 50% of children with high-risk neuroblastoma relapse, often due to drug resistance following induction chemotherapy.

To investigate drivers of chemoresistance in neuroblastoma, we established a longitudinal drug resistance model. IMR-32 cells were pulsed with increasing concentrations of cisplatin over 13 months. Three increasingly resistant cell lines were established at separate time points. We used the 10X Genomics Chromium platform to transcriptionally profile single cells across this resistance trajectory.

We examined gene expression signatures of individual cells to 1) distinguish unique transcriptional cell states, 2) reconstruct transcriptional trajectories, and 3) infer changes in copy number variations (CNVs).

We identified three transcriptionally distinct clusters with varying abundances of cells from all four cell lines. Depending on the cluster assessed, we identified depletion or expansion of transcriptomic phenotypes in response to cisplatin, suggesting that transcriptional plasticity contributes to drug resistance. Through trajectory analysis we observed transitions in cytoskeletal genes with increasing cisplatin exposure. We further linked these changes in gene expression to inferred segmental CNVs found at Chr1p36.

Taken together our work suggests that upon cisplatin exposure, neuroblastoma cells become resistant using a combination of both genetic (innate) and non-genetic (acquired) mechanisms.

SOMETHING'S FISHY ABOUT MY DATA: CHALLENGES WITH REPLICABILITY IN RNA-SEQ

Lachlan Baer, The University of Adelaide

Replicability of results obtained from comparable RNA-seq datasets is largely sought after, however, is not consistently achieved. An initial comparison between two analogous RNA-seq datasets from zebrafish brain tissue identified that substantial batch effects were involved. Further investigation into differences seen between multiple datasets determined a number of factors to contribute to overall variability. Possible sources of variation may occur during the sample and library preparation stages prior to RNA-seq. This suggests that the choice in preceding techniques may have an impact on experimental outcomes, with the potential to obscure particular observations when biological signals

are subtle. An attempt to remove unwanted variation was limited in success. However, consistent results between datasets when analysed for global trends in expression patterns suggests that technique-dependent biological interpretations may still be possible.

SINGLE-CELL RNA-SEQ ANALYSIS TO EXPLORE BONE-MARROW IMMUNE LANDSCAPE

Gunjan Dixit, Australian National University

Bone marrow (BM) contains multiple immune cell subsets with critical functions and is considered an immune regulatory organ. It contains osteoclasts and immune cells fundamentally involved in physiological and pathological bone remodelling. In autoimmune diseases, inflammation can impair the BM niche, disturb hematopoietic and immune development, and induce osteoporosis. Specific cytokines exhibit pleiotropic effects on the immune system, and their discovery in the regulation of survival, differentiation and propagation of activated T cells paved the path for its direct clinical implications in immunotherapy. This project addresses the fundamental question of how low-dose of CytokineX modulates BM immune landscape by comprehensively mapping the therapy-induced changes using single-cell technologies. I analyzed the scRNA-seq data obtained from BM of CD45 cells of mice to identify different immune cell types and compared their expression across four experimental conditions- a control (sham), control treated with CytokineX (Sham+Treatment), ovariectomy-induced osteoporosis (OVX) and OVX treated with CytokineX (OVX+Treatment). The analysis pipeline included quality control, normalization, data integration, dimensionality reduction, clustering and downstream processing. Pairwise differential expression analysis confirmed the identified cell-types and their assignment into clusters. Trajectory analysis using RNA velocity revealed the cellular dynamics and the lineage of sub-populations.

PROFILING GENE DYSREGULATION CAUSING INHERITED PERIPHERAL NEUROPATHY BY CHARACTERISING OF TOPOLOGICAL ASSOCIATED DOMAINS (TADS) IN 3D GENOME

Dr Kaitao, Lai, The University of Sydney

Mr Anthony N.Cutrupi, The University of Sydney

Dr Gonzalo Perez-Siles, The University of Sydney

Professor Garth A. Nicholson, The University of Sydney

Professor Marina L. Kennerson, The University of Sydney

Inherited peripheral neuropathies (IPNs) cause degeneration of peripheral nerves with more than 100 causative genes reported to date. Our studies have demonstrated that how SV can impact on the 3D genome of IPN patients.

We have identified a 1.35 Mb insertion of chromosome 7q36.3 causing gene dysregulation in a large multi-generation family linked to the distal hereditary motor neuropathy DHMN1 locus. We suppose that the DHMN1 complex insertion disrupts genomic organisation leading to gene dysregulation and subsequent axonal degeneration.

We performed high-throughput chromosome conformation capture (Hi-C) analysis of disease-associated mutations and chromosomal rearrangements using DHMN1 induced pluripotent stem cell derived motor neurons (iPSC-MNs) to investigate the mechanism of chromatin interaction underlying the gene dysregulation.

In our works, the 2D contact heatmaps have been generated and the topologically associated domain (TAD) data have been predicted in the view of 3D genome. The neo-TAD on the DHMN1 locus has been identified, which presents duplication and insertion in 3D genome level and may suggest overlapping duplications that extend over the next boundary into the neighbouring regulatory domain. In addition, the 3D representations of the above 2D heatmaps have been constructed, which indicate that the DHMN1 complex insertion has altered the 3D chromatin loops.

THE DYNAMIC GENOME BEHIND THE EMERGENCE OF RECENT OCTOPOD NOVELTIES

Brooke Whitelaw, James Cook University

Cephalopods are characterized by many organismal novelties. To reveal the genomic correlates of organismal novelties, we conducted a comparative study of three octopod genomes. Among the species examined is a member of the blue ringed octopus genus (Hapalochlaena) the only known octopods to store large quantities of the potent neurotoxin tetrodotoxin (TTX) within their tissues and venom gland. We present the first genome of a member of this genus, the southern blue-ringed octopus (Hapalochlaena maculosa) and reveal highly dynamic genome evolution at both non-coding and coding organizational levels. We demonstrate expansions of zinc finger and cadherin gene families associated with neural functions/tissues in both H. maculosa and C. minor are congruent with the previously observations in O. bimaculoides, suggesting an octopod specific trait. Examination of tissue specific genes in the posterior salivary/venom gland (PSG) revealed putative venom proteins, serine proteases dominate expression in O. bimaculoides and C. minor, while representing a minor component in H. maculosa. Voltage-gated sodium channels (Nav) in H. maculosa contain a resistance mutation previously documented in pufferfish and garter snakes to confer 10-15 fold resistance to TTX. No known resistance mutations were identified in either O. bimaculoides or C. minor .

WHAT EXACTLY DO WE KNOW ABOUT FUNCTIONAL CONSTRAINT ON RENAL GENES ASSOCIATED WITH ESRF?

Hope Tanudisastro, The University of Sydney
Mahnoor Bakhtiar, The University of Sydney
Dr Yuan Min Wang, The University of Sydney
Dr Geoff Zhang, The University of Sydney
Dr Hugh McCarthy, The University of Sydney
Professor Stephen Alexander, The University of Sydney

Large-scale exome sequencing data have allowed for the interrogation of low frequency variations and their predicted pathogenicity. It is hypothesised that, in certain genes, selective constraint diminishes observed functional variation.

Using data from the Exome Aggregation Consortium of 60,706 patients, we examined functional constraint on genes associated with end stage renal failure (ESRF) by comparing the frequency of synonymous, missense, and loss-of-function (LoF) mutations against their respective selection-neutral expected values, taking into account gene length, read depth, and local sequence context. ESRF-associated genes were identified from clinical data on paediatric patients in Boston Children's Hospital and the Children's Hospital at Westmead. We compared these values across genes identified by transcriptomic analysis to be highly tissue-enriched in kidneys and those associated with chronic kidney disease (CKD) by genome-wide association study.

Stronger negative selection was observed in ESRF-associated genes than in kidney-expressed genes across LoF mutations ($p<6.6e-07$). Amongst missense mutations, autosomal dominant ESRF-associated genes are under more selective constraint than kidney-expressed genes ($p<0.004$). CKD-associated, ESRF-associated, and kidney-expressed gene sets had similar z-score distributions for synonymous mutations (p = n.s.). Selective pressure was measured most strongly across LoF mutations while genes highly enriched for renal tissue experience relatively minimal negative selection.

MULTI MODAL MACHINE LEARNING TO INFORM BETTER DIAGNOSES, PROGRESSION PROGNOSES AND CLINICAL TRIALS FOR PARKINSON'S DISEASE

Michael Allwright, The University of Sydney

I am employing cutting edge machine learning/Artificial Intelligence algorithms and data management processes to interrogate and bring together multi-modal datasets relating to Parkinson's Disease (PD). Through the integration of these data sources and the application of machine learning methods, I am seeking to develop improved biomarkers thus enabling an improved prediction of disease diagnosis, disease sub-type and disease progression.

These biomarkers will in turn enable patients to receive better treatment plans, bespoke to their specific disease sub-type and rate of progression. It will also enable improved efficiency in clinical trials, due to improved targeting of patient cohorts by disease stage and type.

The data sources considered are:

Publicly available resources such as Michael J Fox/PPMI data and next generation sequencing data as well as clinical data available at Sydney University's Brain and Mind Institute;

Medical Imaging Data, Whole Genome Sequencing Data, Lipidomics, Blood and Metabolonic Data.

The goal is to develop an automated, multi-modal Big Data and machine learning pipeline using the latest technologies (R, Kallisto, Cromwell etc.), which is reusable and scalable and which can be implemented on any relevant data sources to achieve the goals highlighted above.

FORMATION OF BOUNDARIES IN THE DEVELOPING EMBRYO

Bin Wang, Monash University

The heart is one of the most crucial organs throughout life. Its development arises from the embryonic mesoderm and this process is strictly regulated by tightly controlled, spatio-temporal gene expression. Currently, the specific expression network which governs cardiac expression boundaries remains largely cryptic. Using the Tomo-seq database generated by Junker et al. in 2014, of the developing Danio rerio embryo, gene expression can be specifically attributed to 3D spatial regions. By combining known cardiac markers with a powerful computational approach, it is possible to reconstruct detailed gene regulatory networks of the developing heart. Clustering methods can divide similarly expressed genes into groups whereby gaining new biological knowledge following annotation. The hierarchical clustering algorithm is applied to identify and characterise unannotated genes in the LPM domain along the left to right, and anterior to posterior axes. In this way, a comprehensive series of genes active in the LPM territory can be used to inform the reconstruction of a computational GRN, providing new insights into the complex development of cardiac boundaries. In the future, further characterisation of these candidate genes controlling cardiac boundary networks may provide useful insights in clinical intervention in congenital and adult heart disease.

THE ROLE OF KINASES IN THE DIFFERENTIATION OF BONE MARROW STROMAL CELLS INTO OSTEOBLASTS: A SYSTEMATIC ANALYSIS BY KNOCKDOWN

Angelita Liang The University of New South Wales

Kinases may play positive or negative roles in the transduction of key signalling pathways which regulate the differentiation of human bone marrow stromal cells (hBMSCs) into osteoblasts. To discover novel kinases which regulate hBMSC differentiation into osteoblasts, we analysed the data from a systematic RNA interference knockdown of 719 genes including 393 human protein kinases in the hMSC-TERT4 cell line induced to differentiate over 6 days using alkaline phosphatase as a proxy for the extent of differentiation. We have then cross-referenced the putative regulators we found with their mRNA expression profiles over 12 days as measured by time-series RNA Sequencing. By combining knockdown data with large-scale CRISPR knockout data that indicated the essentiality of a gene for cell viability, we identified the genes JAK1,GSK3B,ERBB2 and EPHA3 to be putative positive regulators and PRKCE,PTK7,TEC,PRPS1,PRPS2,PRPSAP1 and PRPSAP2 to be putative negative regulators. We found genes encoding members of the MAPK,MAP2K,MAP3K,MAP4K and NIMA-related kinase families comprising the positive regulator hits, and genes encoding members of the Casein kinase, Protein kinase G, and RIPK families in the negative regulator hits...

FROM DPP4 SUBSTRATES TO BEYOND

Robert Qiao, Flinders University

From type II diabetics, rheumatoid arthritis to Alzheimer, dementia even cancers, chronic diseases create severe social-economical burdens on the healthcare system and patients' welfare are compromised in the wake of prolonged drug usage in many cases.

With ever-increasing omics data becomes available, we now have a unique opportunity to study the global interactions and correlation network of a given enzyme inhibitor therapy. This study attempted to reveal the global interaction network for dipeptidyl dipeptidase 4 (DPP4), and to further explore the possible long-term adverse effects due to prolonged inhibition of DPP4 activity, given that DPP4 has emerged as a new therapeutic target for the treatment of type II diabetics in clinics. The result of this study has obtained strong convergence from both data mining approach and model predictions based on machine learnings, which has depicted a much broader and diverse interaction network then literature suggests.

COMPOSITE SELECTION SIGNALS IN PUREBRED DOGS

Victor Wei Tse Hsu, The University of Sydney

Composite selection signals (CSS) have demonstrated that the locus or interesting regions can be localized in multi-breed populations. Here, we investigated the application of CSS to canine SNP data to assess previously known signatures or identify novel regions from various purebred dogs breed comparisons. We tested the use of CSS to reveal regions associated with canine disease, using lymphoma susceptibility as an example. To gain a more comprehensive insight into lymphoma predisposition, we performed a study using the CSS to analyze selected regions for potential impact on lymphoma incidence. 364 Bullmastiffs were used as a target group with a number of comparative reference groups derived from single or combined breed data. A SNP dataset of gray wolves, previously reported in a European study, was used as a source of ancestral alleles. Using the ancestral or convergent sweeps, clusters of signatures of selection were detected at 101 regions on nine canine autosomes. A gene ontology and pathway analysis of genes in regions identified by CSS revealed 89 candidate genes with enrichment for lymphoma-associated ontologies. The most significant signals were related to the regulation of lymphocyte migration. The CSS is a useful tool for cross-species for identifying potential candidate genes under selection.

BIOINFORMATIC ANALYSIS OF PHOSPHORYLATION-BASED CELLULAR SIGNALLING IN A MODEL OF EPILEPTOGENESIS

Mariella Hurtado Silva, Children's Medical Research Institute

Miss Annika Mayer, The University of Bonn

Professor Susanne Schoch, The University of Bonn

Professor Dirk Dietrich, The University of Bonn

Dr Ashley J. Waardenberg, James Cook University

Dr Mark E. Graham, Children's Medical Research Institute

The application of bioinformatics approaches to various types of â€˜omics data allows greater understanding of biological pathways that are relevant to neurological disease mechanisms. We obtained proteome and phosphoproteome data from a model of temporal lobe epilepsy at 4-hour and 24-hour after the stimuli (an injection with pilocarpine to induce status epilepticus for 30 min). The raw data contained information ono > 40,000 phosphopeptides and was normalized, filtered and reformatted prior to formal analysis. The analysis of the protein phosphorylation data provided information on the most relevant signalling pathways via gene ontology enrichment analysis. The gene ontology terms indicated that phospho-signalling activates transcription factors that respond to strong stimuli and promote neurogenesis. To obtain information on the protein kinases responsible for phospho-signalling, we used KinSwing. KinSwing is a recently developed tool that enables the prediction of protein kinase activity. A number of regulated protein kinases were predicted to be important for each time point and indicate the activation of particular pathways. These bioinformatic analyses enabled insights into how the phosphoproteome is perturbed in the early stages of epileptogenesis.

PHENOTYPE-DRIVEN ANALYSIS OF HUMAN PHOSPHOPROTEOMES

Elise Needham, The University of Sydney

Prioritising the most important phosphosites is a major challenge in phosphoproteomics studies. Often thousands of phosphosites may be regulated and generally only a certain subset of these may drive the biological phenotype of interest. We have developed a method to focus human phosphoproteomics data analysis towards the biologically relevant phenotype. We take advantage of the variability between different humans and directly link this to phenotype. Rather than averaging across humans to compare means for particular treatment groups, we used careful experimental design involving repeated measures within individual humans. We measured the phenotype of interest, in this case glucose uptake, at every time point a sample was taken for phosphoproteomics. With both phenotype and phosphorylation measures across a cohort of humans over 4 different perturbations, we correlated phosphorylation profiles to the phenotype profile. The 152 strongly correlating phosphosites ($r > 0.7$, adjusted p-value < 0.05) out of all 1689 regulated phosphosites (Fold change > 1.5, adjusted p-value < 0.05) were enriched in known glucose uptake regulating phosphosites. The glucose uptake-correlating phosphosites also included novel phosphosites on known glucose uptake-regulating proteins, and potential new regulators. This method extracts the most relevant phosphosites to a phenotype to lead to biological insights.

LEVEL DEPENDENT QBD MODELS FOR THE EVOLUTION OF A FAMILY OF GENE DUPLICATES

Jiahao Diao, University of Tasmania

In our paper, we consider a detailed model with multi-dimensional state-space which consists of binary matrices where rows of a matrix correspond to genes, columns correspond to functions, and the ijth entries record whether or not gene i performs function j. The large state space of this model makes it unsuitable for numerical analysis, but by considering the behaviour of this detailed model we can test the suitability of two alternative models with more tractable state-spaces.

Next, we consider the model proposed in [Teufel 2014], a quasi-birth-and-death process (QBD) with two-dimensional states (n, m). (n, m) records the number n = 1, 2, . . . of genes in the family, and the number m = 0,1,...,n of redundant genes (permitted to be lost). We contrast this to a level-dependent QBD with three-dimensional states (n, m, k) that record additional information k = 1, . . . , K which affects the transition rates.

We show that two-dimensional states (n, m) model is insufficient for meaningful analysis, while the three-dimensional states (n,m,k) model is able to capture the qualitative behaviour of the detailed model. We illustrate the fit between the level-dependent QBD and the original, detailed model, with numerical examples.

TREE SHAPE STATISTICS OF TREES GENERATED USING PHASE TYPE DISTRIBUTED TIMES TO SPECIATION

Albert Christian Soewongsono, University of Tasmania
Associate Professor Barbara Holland, University of Tasmania
Dr Malgorzata O'Reilly, University of Tasmania

This talk will some present preliminary findings in examining tree balance statistics for trees generated using a Coxian Phase Type (PH) distribution of waiting times until speciation. Some earlier results (Hagen et al, 2015) have tried to fit a model that matches with empirical tree data by analysing their tree balance statistics. One of those models is by applying a speciation rate that decreases over species age. This was done by imposing Weibull distribution with shape parameter less than one for speciation time. The biological motivation for using the Weibull was the assumption that a species can be viewed as a collection of i.i.d large populations. The simulation done using that model suggest results that match thousands of empirical trees in terms of their balance. However, viewing those sub-populations as being i.i.d may not be biologically reasonable. Here, we will be using PH distribution, specifically Coxian PH distribution to analyse the problem. The justification in using PH is due to its denseness in the field of all positive-valued distribution and using Coxian PH because every acyclic PH distributions have Coxian PH representation. The early observations using simulations with PH type show a prospective direction towards fitting to empirical tree data.

AN OMICS TRIANGLE: A CASE STUDY OF TRNA GUANINE AND INOSINE-N1-METHYLTRANSFERASE TRM5 IN ARABIDOPSIS THALIANA TO INVESTIGATE THE IMPORTANCE OF TRNA MODIFICATIONS USING TRNA-SEQ, RNA-SEQ, AND PROTEOMICS

Pei Qin (Sabrina) Ng, The University of Adelaide

Transfer RNAs (tRNAs) are critical players in messenger RNA (mRNA) decoding and are often chemically modified. It is essential to understand the consequences of losing tRNA modification by studying tRNA base changes at single-nucleotide resolution. Bioinformatics analysis of tRNA-seq has a higher sensitivity towards less abundant tRNA isotypes. Hence, the combination of tRNA-seq, RNA-seq, and proteomics data analysis enable us to study the molecular consequences of tRNA modification loss. Here, we present a case study of tRNA modifications N1-methylguanosine (m1G) and N1-methylinosine (m1I) at tRNA anticodon loop position 37 (tRNA37), which is essential to maintain translational fidelity by preventing translational frameshift. In Arabidopsis thaliana, AtTRM5, a tRNA Guanine and Inosine-N1-methyltransferase, modifies tRNA37. We show that Attrm5 mutant plants lose m1G and m1I at position 37 in tRNA-Ala and tRNA-Asp using tRNA-seq. Attrm5 mutant plants have overall slower growth and reduced primary root length. Hence, we performed RNA-seq and proteomics data analysis on both wild type and Attrm5 mutant Arabidopsis thaliana plants to examine the changes in gene expression and protein levels. In summary, our triomics approach allows us to gain a greater understanding of how tRNA modification loss affects gene and protein expression, thus impacting plant growth and development.

INTEGRATION OF 'OMICS TECHNIQUES IDENTIFIES EXTENSIVE MITOCHONDRIAL BIOGENESIS AFTER ENDURANCE TRAINING OF HUMAN SKELETAL MUSCLE.

Dr Nikeisha Caruana, The University of Melbourne

In addition to generating the bulk of cellular energy, mitochondria direct a vast array of biological functions essential for cellular homeostasis. A long-standing question in biology concerns the biogenesis of mitochondria and its regulation in response to stress and the metabolic needs of the cellular environment, with exercise representing a major challenge to both these pathways. In order to further demonstrate the effects exercise has on the mitochondria, ten participants underwent three different training volume phases over 12 weeks. Tissue biopsies were taken prior to commencing the study and after each phase, with each mitochondrial isolated proteome analysed by label-free quantitative mass-spectrometry. Proteomics was then integrated with RNA sequencing from muscle biopsies in order to identify trends within both datasets. We observed extensive mitochondrial biogenesis in response to changing volumes of exercise training. While this was met with an overall increase in oxidative capacity, mitochondria underwent extensive remodelling of energetic pathways. Cessation of high-volume exercise reversed some, but not all of these changes. Our findings suggest that training volume is an important determinant of changes in mitochondrial content and function and is a useful model to help to further our understanding of the fundamental mechanisms of mitochondrial biogenesis.

RNA-SEQ ANALYSIS IN A ZEBRAFISH MODEL OF ALZHEIMER'S DISEASE HIGHLIGHTS THE IMPORTANCE OF IRON HOMEOSTASIS

Nhi Hin, The University of Adelaide

Analysing gene expression data from diseased and normal brain tissue has been valuable for exploring molecular mechanisms contributing to Alzheimer's disease and how these diverge from normal aging. Our laboratory has used gene editing technologies to introduce familial Alzheimer's-like mutations into zebrafish, followed by RNA-sequencing of wild-type and mutant brains at young and old age under both normal and low-oxygen conditions in a full-factorial design. This design has offered us a unique opportunity to study the molecular basis of the disease in its early stages in the young brains, and augment our knowledge with how aging and brain oxygen levels relate to disease progression. This poster describes exploratory analyses from RNA-seq data from the brains of these zebrafish and how they allow us to explore broad-scale disruptions in molecular processes, in addition to more detailed analyses testing whether processes hypothesised to be important in Alzheimer's disease (iron homeostasis in particular) were disrupted at the regulatory level. An important theme of the poster is the importance of data visualisation in facilitating hypothesis generation and communicating findings in an accessible way.

DRUGS MODULATING STOCHASTIC GENE EXPRESSION AFFECT THE ERYTHROID DIFFERENTIATION PROCESS

Dr Anissa Guillemin, The University of Melbourne

To understand how a metazoan cell takes the decision to differentiate, we study the role of stochastic gene expression during the erythroid differentiation process. It has been settled that SGE participate in decision making at the single cell level. However, experimental evidence of the relation between is still lacking. Using single cell transcriptomic analyses on avian erythropoiesis, we selected 3 drugs able to modulate the level of SGE.

We then assessed if these drugs that modulate SGE can also affect the differentiation process. We show that drugs reducing the SGE amount, significantly decreased the percent of differentiated cells and inversely.

We used a mathematical model to estimate which parameters were modified by drug treatment. We observed that among the affected parameters of the model, the rate of differentiated cells remains the parameter the most strongly affected by all drugs, supporting the previous results.

Therefore, using single-cell analyses and modeling, we provide the first evidence for a positive relation between SGE level and cell differentiation, leading to a new potential way to control this process.

TRACKING LEUKOCYTES IN INTRAVITAL TIME LAPSE IMAGES USING 3D CELL ASSOCIATION LEARNING NETWORK

Marzieh Rahmani Moghadam, La Trobe University

Leukocytes are key cellular elements of innate immune system in all vertebrates, which play a crucial role in defending organism against invading pathogens. Tracking these highly migratory and amorphous cells in in vivo models such as zebrafish embryos is a challenging task in cellular immunology. As temporal and special analysis of these imaging datasets by human operator is quite laborious, developing automated cell tracking method is highly in demand. Despite the remarkable advances in the cell detection, this field still lacks powerful algorithms to accurately associate the detected cell across time frames. The cell association challenge is mostly related to the amorphous nature of cells, and their complicated motion profile through their migratory paths. To tackle the cell association challenge, we proposed a novel deep-learning-based object linkage method. For this aim, we trained our proposed 3D cell association learning network (3D-CALN) with enough manually labelled paired 3D images of single fluorescent zebrafish's neutrophils from every two consecutive frames. A comparison of our tracking accuracy with other available tracking algorithms shows that our approach performs well in relation to addressing cell tracking problems.