1

CRISPR-Cas Gene Editing: Biology, Technology and Ethics

Jennifer Doudna[1]

1. UC Berkeley/Howard Hughes Medical Institute, Berkeley, CALIFORNIA, United States

Gene editing with CRISPR technology is transforming biology. Understanding the underlying chemical mechanisms of RNA-guided DNA and RNA cleavage provides a foundation for both conceptual advances and technology development. I will discuss how bacterial CRISPR adaptive immune systems inspire creation of powerful genome engineering tools, enabling advances in both fundamental biology and applications in medicine. I will also discuss the ethical challenges of some of these applications.

2

Principles of epigenetics and chromatin in development and human disease

Ali Shilatifard[1]

1. Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

Epigenetic regulation of gene expression in metazoans is central for establishing cellular diversity, and the perturbation of this process results in pathological conditions. Although transcription factors are essential for implementing gene expression programs, they do not function in isolation and require the recruitment of various chromatin-modifying and remodeling machineries. A classic example of developmental gene expression through chromatin is the regulation of the balanced activities of the Polycomb group (PcG) proteins within the PRC1 and PRC2 complexes, and the Trithorax group (TrxG) proteins within the COMPASS family. Recent large scale genome sequencing efforts of human cancer have demonstrated that PcG and COMPASS subunits are highly mutated in a large number of human solid tumors and hematological malignancies. I will discuss our laboratory's latest biochemical and genetic studies defining the molecular properties of COMPASS and PcG families in the regulation of gene expression, during development, the central role they play in cancer pathogenesis, and how we have taken advantage of such basic molecular information to develop targeted therapeutics for the treatment of hematological malignancies, pediatric brain cancer, and other forms of solid tumors.

3

Defining how problematic DNA replication impacts on chromosome segregation

Ian D Hickson[1], **Rahul Bhowmick**[1], **Anna H Bizard**[1], **Ozgun Ozer**[1], **Manikandan Paramasivam**[1], **Kata E Sarlos**[1], **Roshan Singh Thakur**[1], **Andres Bueno Venegas**[1]

1. Centre for Chromosome Stability, Department of Cellular and Molecular Medicine, University of Copenhagen, Copenhagen, Denmark

During tumorigenesis, the activation of oncogenes generates DNA replication stress, which leads to replication-associated genome rearrangements. This stress is not felt equally across the genome, but instead is concentrated at certain 'difficult-to-replicate' loci. The most prominent of these are multiple conserved loci called common fragile sites (CFSs). It was while studying CFSs that we identified an unusual DNA structure that had escaped detection previously despite being present in virtually all human cell anaphases; the ultra-fine anaphase bridge (UFB). UFBs are thin threads of DNA that connect the separating sister DNA masses in anaphase. They cannot be stained with DNA dyes and do not contain histones, making their detection problematic and dependent on immunofluorescence for associated proteins such as PICH or BLM. UFBs arise from specific loci that are characterized by the unusual structure and replication program; most notably, centromeres, CFSs, the rDNA, telomeres. We are conducting a detailed analysis of two aspects of UFB biology: modeling the association of proteins to UFBs *in vitro* using optical tweezers, and investigating how unresolved UFBs affect cell division. In addition, through analysis of the underlying basis of UFB formation in human cells, we identified that CFSs and telomeres delay the completion of their replication program until early mitosis - in a process that we term MiDAS (mitotic DNA synthesis). MiDAS depends upon a subset of homologous recombination factors, such as RAD52, and appears to be a form of break-induced replication previously characterized in detail only in yeast. Inhibition of MiDAS leads to major chromosome segregation abnormalities. The latest progress on these projects will be presented.

4

Imaging chromatin dynamics during the DNA damage response by FLIM-FRET microscopy

Elizabeth Hinde[1, 2], Jieqiong Lou[1], Belinda Wright[2], Tara K Bartolek[3], Tony Cesare[3]

1. Biochemistry and Molecular Biology Department, University of Melbourne, Melbourne, VIC, Australia

2. EMBL Australia node in Single Molecule Science, University of New South Wales, Sydney, NSW, Australia

3. Genomic Integrity Group, Children's Medical Research Institute, Westmead, NSW, Australia

Chromatin dynamics modulate DNA repair factor accessibility throughout the DNA damage response. The spatiotemporal scale upon which these dynamics occur render them invisible to live cell imaging. Here we employ fluorescence lifetime imaging microscopy (FLIM) for FRET detection of nucleosome arrangement in live cells and monitor the structural rearrangements of chromatin during DNA repair. With this technology we demonstrate that genomic double strand breaks induce both local and global condensation events in the chromatin network and the detected chromatin dynamics facilitate DNA repair factor recruitment specifically to the lesion site.

5

Actin polymerization alters nuclear architecture in response to DNA replication stress to maintain genome stability

Anthony J Cesare[1], Noa Lamm-Shalem[1]

1. Children's Medical Research Institute, Westmead, NSW, Australia

Impediments that that slow the rate of DNA replication are collectively referred to as "replication stress". Replication stress is the main driver of genome instability in early cancer development and is recognized as a hallmark of cancer. Actin is a cytoskeletal protein that forms filaments to provide cells with mechanical support and driving force for movement. While actin is traditionally considered a cytoplasmic protein, recent evidence indicates actin polymerization can also occur inside the nucleus. However, the role for nuclear actin fibres, the mechanism(s) triggering their polymerization, and the impact of nuclear actin on the genome remains unclear.

Using live-cell and super-resolution imaging, chromatin fibre analysis, biochemistry, cell and molecular biology, we discovered that actin polymerization plays a prominent role in the replication stress response. Consistent with induced DNA replication stress, pharmacological inhibition of actin polymerization in human cells resulted in S-phase elongation, reduced DNA replication rate, shortened distance between replication origins, and increased occurrence of micronuclei and anaphase abnormalities. Pharmacological replication stress induced ATR and mTOR-dependent nuclear actin polymerization, which altered the nuclear architecture through the expansion of nuclear volume and redistribution of replication forks to the nuclear periphery. Inhibiting ATR, mTOR or actin polymerization, suppressed replication stress-dependent nuclear alteration and prevented the restart of stalled replication forks. Preliminary data indicate co-localization of stalled replication forks with certain repair factors is compartmentalized to the nuclear periphery, suggesting that actin polymerization facilitates movement of stalled forks to the nuclear periphery for repair.

Cumulatively, these data reveal a novel pathway where actin dependent forces shape the nucleus in response to replication stress to maintain genome health. This findings suggest nuclear actin polymerization may have additional nuclear roles that that impact genome function

11

Editing wild populations: local gene drive, evolutionary stability, and community guidance

Kevin Esvelt[1]

1. Massachusetts Institute of Technology, MA, USA, United States

Gene drive systems could address many health and environmental problems by altering or suppressing wild populations, but the possibility that self-propagating gene drives will spread across international borders is likely to preclude most applications. Localized gene drive systems are more promising, yet must be empirically demonstrated to be evolutionarily stable. Given that most organisms cannot be raised in very large numbers, might other species be used as proxies? Finally, because people cannot opt-out of ecological changes, conducting traditional closeted research on gene drive applications effectively denies people a voice in decisions intended to affect them. Inviting community guidance of research may increase safety, efficacy, and the likelihood of eventual public support for applications.

12

Conservation of transcriptional variation across human, mouse and … armadillo?!

Sara Ballouz[1], Jesse Gillis[1]

1. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, United States

Controlling for sources of variability within a targeted experiment is standard practice in gene expression studies, where transcriptional noise is inherent and not fully understood. A common control is to condition on genetic background, usually through inbred model organisms, strains, clones or cell lines. Although this decreases variability within the experimental system, generalizing to another genetic background or strain is perilous. This is rarely more true than in translational or comparative biology, where genetic background changes profoundly (i.e., across species).

In order to assess the impact of genetic background on mammalian transcriptomic findings, we exploited the polyembrony of the wild nine-banded armadillo (*Dasypus novemcinctus*). It is an ideal model system for our purposes since it is outbred, yet produces monozygotic quadruplets in every litter, serving as natural biological replicates. Surprisingly, this unique reproductive mode has yet to be exploited in transcriptional studies, even though its discovery well over a hundred years ago was critical to the field of developmental biology.

First, we sequenced the blood transcriptomes of five litters of armadillo quadruplets to generally assess transcriptional variation. We found 2982 genes with human and mouse homologs exhibiting statistically significant differences (FDR<0.001) between quadruplet sets, indicating variability sensitive to genetic background. Immune function and cell cycle homologs were particularly prominent. To determine if these genes, sensitive to genetic background, were generating spurious results within mouse and human experiments, we performed a meta-analysis across 3275 pre-existing gene expression studies. We find highly variable genes are often called differentially expressed in mouse (rho=0.28, p<2.2e-16) and human (rho=0.27, p<2.2e-16).

Our findings suggest that genes sensitive to genetic background can be easily identified and are a potentially useful probe for results that will not generalize across species, helping to address the replicability “crisis” in transcriptomic, functional genomics and beyond.

13

Accounting for systematic bias in bulk and single cell RNA-Seq data

Rafael Irizarry[1]

1. Dana-Farber Cancer Institute, Harvard University, Boston, MA, United States

In this talk I will demonstrate the presence of bias, systematic error and unwanted variability in next generation sequencing. I will show the substantial effects these have on downstream results and how they can lead to misleading biological conclusions. I will do this using data from the public repositories as well as our own. We will then describe some preliminary solutions to these problems.

14

Single-cell tumor transcriptomics: algorithms and applications

Shyam Prabhakar[1]

*1. A*STAR Genome Institute of Singapore, Singapore, 138672*

Single-cell RNA-seq (scRNA-seq) can in principle reveal the diversity of cell types or functional states within a heterogeneous tumor sample. However, major improvements are needed in sample processing and data analytics to achieve this goal. We have developed a suite of algorithms that are computationally inexpensive, make minimal statistical assumptions and comprehensively outperform existing approaches when benchmarked on scRNA-seq data. Based on these algorithms and a custom tissue dissociation protocol, we analyzed single-cell transcriptomes from >30 unsorted colorectal and lung primary tumors, with matched normal mucosa. The analysis yielded novel insights into tumor-specific stromal signatures, epithelial mesenchymal transition (EMT), cancer cell stemness and patient survival, and also revealed multiple cancer-associated fibroblast subtypes. Overall, our results indicate that hypothesis-free tumor-vs-control scRNA-seq of unsorted cell populations could be used as a general approach for investigating cancer mechanisms.

15

Transcriptional regulation of neuronal function by the histone demethylase KDM5

Julie Secombe[1]

1. Einstein College of Medicine, Clifton, NEW JERSEY, United States

My lab is interested in understanding the functions of the KDM5 (lysine demethylase 5) family of transcriptional regulators that act by binding to, and enzymatically modifying, chromatin. Specifically, I will be talking about our efforts to define how mutations in *kdm5* genes cause heritable forms of syndromic and non-syndromic intellectual disability in humans. To do this, we are utilizing the genetic toolkit available by using the model organism *Drosophila*. In addition to examining the effects of loss of KDM5 function using a null mutation, we are leveraging the fact that all 19 disease-associated missense mutations affect evolutionarily conserved residues. These mutations do not cluster in a particular domain of KDM5, thus it is unknown whether they alter similar or distinct target genes and pathways to affect cognition. To date, we have generated nine fly strains harboring intellectual disability-associated missense mutations and have begun defining the transcriptional, behavioral and cellular defects of these alleles. Our data reveal striking learning and memory defects that occur with or without altering neuronal morphology. Transcriptome analyses of these alleles also identify a cohort of overlapping genes that are likely to be key to the relationship between KDM5 and neuronal (dys)function.

16

Chromatin modifiers SET-32 and SET-25 establish a transgenerational silencing signal in *Caenorhabditis elegans*

Rachel Woodhouse[1], **Alyson Ashe**[1]

1. University of Sydney, Sydney, NSW, Australia

Mounting evidence in several organisms suggests that some epigenetic modifications acquired by an individual during its lifetime can be inherited for multiple future generations. This phenomenon, termed transgenerational epigenetic inheritance, may provide a mechanism for the inheritance of environmentally acquired traits. We are studying transgenerational epigenetic inheritance using a model organism, the nematode *Caenorhabditis elegans*.

We have developed a system in which RNAi-induced silencing of a GFP transgene is robustly inherited for multiple generations in the absence of the initial RNAi trigger. We show that the histone methyltransferase SET-25 and the putative histone methyltransferase SET-32 are required for effective transmission of transgene silencing. Specifically, whilst *set-25* and *set-32* mutant animals exposed to RNAi display transgene silencing, their unexposed offspring fail to inherit silencing. Intriguingly, the few animals which escape this failure and remain silenced then produce subsequent generations of silenced progeny. Furthermore, *set-25* and *set-32* mutants segregated from silenced *set-25*/+ and *set-32*/+ heterozygotes respectively remain fully silenced. Together, this data suggests that SET-25 and SET-32 are required for the establishment of a transgenerational silencing signal, but not for long-term maintenance of this signal between subsequent generations. We thus propose a three-step model of transgenerational epigenetic inheritance consisting of Initiation, Establishment and Maintenance. Small RNA sequencing experiments are underway to further test this.

SET-32 is uncharacterised, so we are also investigating its broader functions. We show that *set-32* mutant hermaphrodites have reduced fertility, producing fewer live offspring, fewer total eggs and more unfertilised eggs in comparison to wild-type animals. This reduced fertility is rescued by providing mutant hermaphrodites with wild-type sperm, indicating a male germline defect. Closer investigation reveals that *set-32* mutant sperm are defective in crawling. Furthermore, *set-32* mutant animals display extended lifespan independent of reduced fertility, providing new insight into the little-understood contribution of chromatin modifiers to lifespan regulation.

17

DNA REPLICATION TIMING SHAPES THE CANCER EPIGENOME

Qian Du[1], **Saul A Bert**[1], **Nicola J Armstrong**[2], **Elizabeth Caldon**[3], **Clare Stirzaker**[4, 1], **Susan Clark**[4, 1]

1. Genomics and Epigenetics, Garvan Institute of Medical Research, Sydney, NSW, Australia

2. Mathematics and Statistics, School of Engineering and Information Technology, Murdoch University, Murdoch, Western Australia, Australia

3. Cancer Division, Garvan Institute of Medical Research, Sydney, NSW, Australia

4. St Vincent's Clinical School, University of NSW, Sydney, New South Wales, Australia

Cancer cells are highly proliferative and undergo successive rounds of DNA replication. Replication of the genome follows a highly organised 'replication-timing' program where genomic domains are replicated in a specific temporal order during S-phase, from early to late. Replication timing is known to modulate the faithfulness of DNA replication. As the transmission of epigenetic information during S- phase is intimately associated with the replication fork, replication timing may similarly modulate the faithful transmission of the epigenome. Here, we investigate how the replication timing program organizes the epigenome and contributes to aberrant epigenetic change in normal prostate and cancer cells.

We performed Repli-Seq that utilizes next generation sequencing to map the order of nascent DNA replication across S-phase in normal and prostate cancer cells. We find that, while replication timing is mostly maintained from normal to cancer, domains of replication timing change are accompanied by aberrant changes in gene expression, chromatin marks and DNA methylation. Strikingly, domains of replication timing change correspond to our previously identified domains of Long Range Epigenetic Silencing and Activation found in prostate cancer (LRES, LREA). Importantly, late replication appears to predispose cancer cells to aberrant changes in heterochromatin and DNA methylation, and becomes enriched for chromosomal rearrangements in cancer. Furthermore, we find that different replication times biases towards either cis or trans chromosomal rearrangements. We propose that this mutational bias is related to the differences in cancer epigenetic remodelling that occur in early, compared to late, replication timing. Our findings highlight for the first time that the replication timing program associates with aberrant epigenetic remodelling in cancer, and together, potentially contributes to the mutational pattern in cancer.

18

3'-deoxyadenosine induces bulk transcript lengthening via alternative polyadenylation

Belinda J Goldie[2, 1], Rachael E Turner[2, 1], Paul F Harrison[3], Angavai Swaminathan[2, 1], David R Powell[3], Traude H Beilharz[2, 1]

1. Dept of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria, Australia

2. Monash Biomedicine Discovery Institute, Clayton, VICTORIA, Australia

3. Monash Bioinformatics Platform, Monash University, Clayton, Victoria, Australia

Key regulatory elements that influence the translational fate of mRNA, such as RNA binding proteins, microRNA and m^{6}A methylation, display strong site preference for motifs positioned in the 3'-UTR. In eukaryotes, up to 80% of mRNAs regulate 3'-UTR length by encoding multiple locations for cleavage and polyadenylation (CPA) of the nascent transcript, known as alternative polyadenylation (APA). 3'-deoxyadenosine (3'-dA, also called cordycepin) is a naturally-occurring adenosine analogue known to cause bulk mRNA shortening via chain termination due to its lack of reactive hydroxyl group at the 3' position. 3'-dA has been used in Chinese medicine for many years for its anti-inflammatory, anti-oxidant and anti-bacterial properties, however its effects on RNA metabolism are poorly understood. Using poly(A)-tail-focused deep sequencing, PAT-seq, we show that, paradoxically, treatment with 3'-dA induces bulk transcript lengthening, favouring distal APA site selection; this phenotype was conserved between yeast and mouse primary bone-marrow-derived macrophages (BMDMs). A PolII slow mutant yeast strain reversed the phenotype, while strains with defective CPA machinery recapitulated it, implicating transcription rate and CPA complex stoichiometry as key determinants of CPA site choice. We also observed signs of mitochondrial dysfunction, and indeed cordycepin-treated BMDMs were unable to respond to LPS challenge, retaining an anti-inflammatory profile due to suppressed interferon-gamma induction. Together our results suggest that the therapeutic effects of 3'-dA lie outside of its established function as an mRNA chain terminator, and this molecule may be applicable to a broader spectrum of inflammatory and transcript-related disorders.

19

Intron retention: A widespread and conserved mechanism of gene expression control regulated by epigenetic changes

Justin JL Wong[1]

1. Centenary Institute, Camperdown, NSW, Australia

Intron retention (IR) is a mode of alternative splicing that occurs when an intron is not excised by the spliceosome, and is preserved in the final mature mRNA. Unlike many other alternative splicing events that promote protein diversity, IR is a distinctive mechanism that regulates genes expression via cytoplasmic nonsense-mediated decay or nuclear-enriched degradation mechanism(s). Herein, I will describe the discovery of IR as a widespread mechanism that controls the expression of functionally-related genes during haemapoietic cell differentiation. I will summarise my previous findings in granulocytes (1), erythroblasts and megakaryocytes, and new finding in monocytes and macrophages. I will present recent findings that IR is regulated by epigenetic changes. In particular, I will highlight my own work that IR is facilitated by

reduced recruitment of splicing factors via decreased DNA methylation levels near splice junctions (2). I will also present data concerning the conservation of IR over 430 years of evolution in vertebrates (3). Finally, I will discuss the roles of IR in human cancers including its potential functions in regulating the expression of tumour suppressor genes and oncogenes.

1. **Wong JJL***, Ritchie W*, Ebner O, Selbach M, Wong JWH, Huang Y, Gao D, Pinello N, Gonzalez M, Baidya K, Thoeng A, Khoo T-L, Bailey CG, Holst J & Rasko JEJ. (2013) Orchestrated intron retention regulates normal granulocyte differentiation. *Cell* 154:583-595.(*Equal first)

2. **Wong JJL***#, Gao D*, Nguyen TV*, Kwok C-T, van Geldermalsen M, Middleton R, Pinello N, Thoeng A, Nagarajah R, Holst J, Ritchie W#, Rasko JEJ#. (2017) Intron retention is regulated by altered MeCP2-mediated splicing factor recruitment. *Nat Commun*. 8:15134.(*Equal first # co-corresponding)

3. Schmitz U, Pinello N, Jia F, Alasmari S, Ritchie W, Keightley MC, Shini S, Lieschke G, **Wong JJL**†, Rasko JEJ†. (2017) Intron retention enhances gene regulatory complexity in vertebrates. *Genome Biol*. 18:216.(†Equal last)

20

Understanding the role of eosinophils in adipose tissue energy expenditure

Kate GR Quinlan[1], Alexander Knights[1], Emily Vohralik[1], Elizabeth Stout[1], Merlin Crossley[1]

1. School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia

Obesity is a global problem and represents a significant health and economic burden. Our research concentrates on understanding the hormones and molecular pathways that drive obesity.

Recently a new category of brown fat-like cells, so-called “beige” cells residing within subcutaneous white fat, has been identified. These cells burn fuels to generate heat and therefore may reduce obesity by burning off rather than storing excess fuels. Cells of the immune system – macrophages, innate lymphoid cells and recently eosinophils – appear to be essential to the “browning” of white fat cells.

While studying mice with a deletion in the gene encoding the transcription factor Kruppel-like Factor 3 (KLF3) we made a number of serendipitous discoveries. These mice are lean, even when fed a high fat diet and have higher numbers of eosinophils within their fat. Interestingly, these mice show evidence of an increased capacity for thermogenesis even when housed at room temperature.

We have performed genome-wide expression analyses on eosinophils isolated from white adipose tissue and saw expression of a number of genes that encode for known browning secreted proteins. Our data suggest that eosinophils may contribute to beige fat biogenesis by secreting these factors. The eosinophils from *KLF3* knockout mice, where we see enhanced browning, expressed higher levels of these secreted proteins. Interestingly, we also detected expression of a number of novel secreted proteins in adipose tissue-derived eosinophils. We are now testing whether these novel secreted proteins are able to induce browning in cell culture and *in vivo* models.

Our data suggest that adipose tissue-resident eosinophils secrete a number of factors to drive the browning of adipocytes. This emphasises the importance of eosinophils in the browning process. Our study of these factors may provide a platform for the development of new therapeutic agents to drive browning and combat obesity.

21

Modelling breast cancer progression using massively-parallel single-cell RNAseq technology.

Fatima Valdes Mora[1, 2], Rob Salomon[1], Brian Gloss[1, 2], Yolanda Colino-Sanguino[1], Daniel Roden[1, 2], Andrew Law[1], Kendelle Murphy[1], James Conway[1], Marcel Dinger[1, 2], Samantha Oakes[1, 2], Paul Timpson[1, 2], Christopher Ormandy[1, 2], David Gallego Ortega[1, 2]

1. Garvan Institute of Medical Research, Darlinghurst, NSW, Australia

2. St. Vincent's Clinical School, Faculty of Medicine, UNSW, Sydney, NSW, Australia

Transcriptome analysis has been extensively used to understand the heterogeneity of breast tumours, defining intrinsic molecular subtypes and signatures able to predict response to therapy and patient outcome. This molecular phenotyping has fostered crucial therapeutic advances. However, cancer cell diversity constitutes a challenge for cancer treatment and deeply impact the outcome of cancer patients. A simultaneous overview of cancer cells and associated stromal cells is critical for the design of improved therapeutic regimes.

Single-cell RNA-seq has emerged as a powerful method to unravel heterogeneity of complex biological systems; this has enabled *in vivo* characterization of cell type compositions through unsupervised sampling and modeling of transcriptional states in single cells. Here we use the high-throughput microfluidic-based single-cell RNA-seq method Drop-seq to elucidate the function and cellular composition of breast tumours. We

use the MMTV-PyMT mouse mammary tumour model to provide high-resolution landscapes of the disease progression, delivering simultaneous observation of cellular events that result in the acquisition of the metastatic phenotype. We define transcriptional networks that result in acquired immune tolerance, extracellular matrix remodelling and progression to metastatic disease. The unprecedented resolution generated by analysis 40,000 individual tumour cells revealed dynamics and plasticity of cancer cells during progression to metastatic disease. Finally, we studied the functional consequences of identified cell signatures of well-known key events during breast cancer progression, EMT, collagen deposition, inflammation and hypoxia.

In summary, we provide a large-scale single-cell transcriptional landscape of breast tumours that allows unprecedented understanding of breast heterogeneity and deep analysis of the events that result in cancer progression. scRNA-seq technology is generating a paradigm-shift in our understanding of biology, applied to tumour biology will lay the first stone for the development of more specific cancer therapies.

22

Genome sequencing of 15,000 healthy elderly Australians

Paul Lacaze[1], Moeen Riaz[1], ASPREE Investigator Group, ASPREE Healthy Ageing Biobank, Robyn Woods[1], John McNeil[1]

1. Monash University, Melbourne, VIC, Australia

The ASPREE Healthy Ageing Biobank contains ~15,000 consented samples from Australians aged 70 years or older participating in the ASPirin in Reducing Events in the Elderly (ASPREE) study - Australia's largest clinical trial and longitudinal study of healthy ageing [1]. At enrolment, all ASPREE participants were confirmed to be free of major life-threatening cardiovascular disease, cancer or cognitive decline, meaning samples were ascertained from confirmed healthy elderly individuals depleted of typical monogenetic disease phenotypes [2]. All ASPREE biobank samples are sequenced using a targeted 'super-panel' of 750 genes used commonly in clinical testing, which includes all ACMG59 clinically actionable genes plus pan-cancer, cardiovascular and neurological gene coverage. Over 9,000 samples have been sequenced so far (Oct 2017), already identifying hundreds of actionable pathogenic variants in individuals lacking any apparent signs and symptoms of genetic disease beyond 70 years [3]. Results will be presented on these findings, with implications for our understanding of penetrance and clinical actionability for genes used in routine testing [4]. Through collaboration, these results will help inform the Resilience Project, a global effort to identify individuals with highly penetrant pathogenic variants who do not appear to develop typical signs and symptoms of disease well beyond the expected age of onset [5, 6]. In addition, ASPREE has conducted whole genome sequencing on 2000 of the oldest, cancer-free Australian participants as part of the Medical Genome Reference Bank (MGRB) project. This presentation will give an overview of the ASPREE trial, depth and breadth of longitudinal phenotype data and fiindgs from genomic research.

23

Identifying the cause and function of T>G mutations in oesophagus and gastric cancer genomes

Jayne A Barbour[1], Rebecca C Poulos[1], Yizhou Huang[1], Julie AI Thoms[1], Dominik Beck[1], John E Pimanda[1], Jason WH Wong[1]

1. UNSW Sydney, Randwick, NSW, Australia

Accumulation of somatic mutations is necessary for oncogenic transformation. Somatic mutation patterns in cancer fall into 30 distinct "signatures" based on combinations of single nucleotide variants in the trinucleotide context. The aim of this project is to explore the mechanisms and aetiology underlying COSMIC Signature 17 (Sig17) and impact on cancer development. Sig17 is present in 80% of eosophageal adenocarcinoma (EAC) and 40% gastric adenocarcinoma (GAC). It is characterised by high frequency of T>G mutations which occurs when oxidised guanine (oxoG) from the dNTP pool erroneously pairs with adenine. Bile reflux is a risk factor for EAC and causes oxidative stress and so we hypothesise bile causes Sig17. Interestingly, Sig17 mutations also form hotspots in motifs of CCCTC-binding factor (CTCF) binding sites which are important for transcriptional repression and DNA loop formation. Therefore, Sig17 mutations may drive cancer through altered DNA conformation and long-range gene expression.

Analysing COSMIC Cell Lines Project data revealed the gastric adenocarcinoma cell line AGS as having strongest Sig17 profile. We performed whole genome sequencing on AGS cells at 30X to confirm a weighted Sig17 profile of 23%. Overlapping our AGS mutation calls with CTCF motifs revealed 34 hetereozygous mutations in CTCF binding sites. We subsequently performed CTCF ChIP-seq with allele-specific analysis to confirm that mutant alleles cause loss of CTCF binding. We are currently performing allele-specific capture Hi-C to investigate whether loss of CTCF binding results in changes to DNA loop formation and plan to integrate these findings with allele-specific RNA-seq data to determine if altered loop formation causes oncogenic gene

expression. To explore how oxo-guanine gets incorporated into DNA we are utilising an oxoG specific antibody to immunoprecipitate DNA followed by sequencing.

Understanding the aetiology and biological consequences of Sig17 is important for development of preventative strategies and targeted drug selection in these cancers.

24

Cardiac directed differentiation using small molecule Wnt modulation at single-cell resolution

Clayton Friedman[1], Quan Nguyen[1], Samuel Lukowski[1], Abbigail Helfer[1], Han Chiu[1], Holly Voges[1], Greg Baillie[1], Anne Senabouth[1], Angelika Christ[1], Timothy Bruxner[1], Charles Murry[2], Emily Wong[1], Yuliang Wang[2], James Hudson[1], Ziv Bar-Joseph[3], Patrick Tam[4], Joseph Powell[1], Nathan Palpant[1]

1. University of Queensland, St. Lucia, QLD, Australia

2. University of Washington, Seattle, WA, USA

3. Carnegie Mellon University, Pittsburgh, PA, USA

4. University of Sydney, Sydney, NSW, Australia

Differentiation into diverse cell lineages requires orchestration of gene regulatory networks guiding diverse cell fate choices. Utilizing human pluripotent stem cells, we measured expression dynamics of 17,718 genes from 43,168 cells across five time points during cardiac-directed differentiation. We used unsupervised clustering to identify the transcriptional phenotype of 15 subpopulations correlating with germ layer and cardiovascular differentiation *in vivo.* These data reveal transcriptional networks underlying lineage derivation of mesoderm, definitive endoderm, vascular endothelium, cardiac precursors, and definitive cardiac fates including contractile cardiomyocytes and non-contractile derivatives. Utilizing a customized lineage trajectory prediction algorithm, *scdiff*, we analyzed transcription factor regulatory networks as a basis to link subpopulations from time course single cell data to predict fate choices from pluripotency into the cardiac lineage. While contractile cardiomyocytes follow a trajectory of known gene regulatory networks, we identified PBX1, a transcription factor involved in outflow tract (OFT) development, as a regulator of non-contractile cardiac derivatives. This phenotype was confirmed by Spearman rank correlations comparing *in vitro*-derived OFT cells against single cell isolated subpopulations of the E9.5 mouse heart. We identified the non-DNA binding homeodomain protein, HOPX, as a candidate regulator underlying cardiomyocyte vs OFT fates *in vitro* and *in vivo*. While HOPX is one of the earliest functional regulators of cardiac fate *in vivo*, we show that HOPX is expressed late and in only 16% of cardiomyocytes and has a repressive chromatin state during *in vitro* differentiation. Using genetic loss of function hiPSCs coupled with physiological assessment of engineered heart tissues, our findings implicate dysregulation of HOPX during *in vitro* cardiac-directed differentiation underlying the molecular and physiological immaturity of stem cell-derived cardiomyocytes. Overall, this study provides the first single cell decomposition of cardiac directed differentiation providing a basis for dissecting *in vitro* cell fates for developmental studies and disease modelling.

25

Essential roles for minor class splicing in development and cancer

Karen Doggett[1], Ben B Williams[1], Sebastian Markmiller[2], Kimberly J Morgan[1], Kate D Sutherland[1], Robert J O'Donoghue[3], Tracy L Putoczki[1], Matthias Ernst[3], Joan K Heath[1]

1. Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia

2. Department of Surgery, Royal Melbourne Hospital, Parkville, Vic, Australia

3. Cancer and Inflammation, Olivia Newton John Cancer Research Institute, Melbourne, Vic, Australia

Splicing is an essential step in eukaryotic gene expression. While a clear majority of introns is excised by major class (or U2-dependent) splicing, the appropriate excision of a very small subset of introns (<0.35% in human) depends on minor class (or U12-dependent) splicing. However, these introns are highly conserved and genes that contain minor class introns are over-represented in functions and pathways related to essential cellular processes, such as cell cycle, DNA replication and repair, RNA processing, and important signalling pathways including RAS/MAPK signalling.

Our characterisation of a zebrafish mutant, *caliban,* lacking a specific minor class spliceosome component, Rnpc3, revealed minor class splicing was vital for vertebrate development. Transcriptome analysis of *rnpc3*$^{-/-}$ larvae demonstrates that minor class splicing is required for the proper expression of genes involved in transcription, splicing and nuclear export. Furthermore, since these genes are essential for the growth and division of rapidly proliferating cells, we hypothesized that efficient minor class splicing would also be crucial for cancer cells. By utilizing a *kras*^{G12V}-driven zebrafish hepatocellular carcinoma model we found that heterozygous loss of *rnpc3* restricted liver hyperplasia.

In addition, we generated germline and conditional mouse models of *Rnpc3* deficiency to investigate the importance of minor class splicing during mammalian development and in tumour-prone mice. Rnpc3 is indispensable for mouse development and systemic loss via inducible deletion in adult mice results in rapid weight loss, leukopenia and atrophy of the gastrointestinal tract. Remarkably, we found that *Rnpc3* heterozygosity in mice decreases gastric and lung tumour burden and also prolongs survival of *Pten* heterozygous mice, which spontaneously develop large lymphomas.

This study presents the first mouse models of minor class splicing deficiency and establishes that this hitherto under-appreciated process is indispensable throughout life and may represent a useful, clinically relevant target for a broad spectrum of cancer types.

26

MicroRNAs act on and within regulatory networks

Greg Goodall[1]

1. Centre for Cancer Biology, Adelaide, SA, Australia

It is becoming evident that microRNAs operate within regulatory networks, involving both cooperative actions of multiple microRNAs as well as the targeting by individual microRNAs of multiple components within regulatory pathways and networks. Investigations of the role of the miR-200 family of microRNAs in controlling epithelial to mesenchymal transition (EMT), which has major influence on the propensity of cancer cells to invade, to metastasise, and to resist cancer therapies, have highlighted the importance of microRNA interactions within regulatory networks. The miR-200 family restricts the expression of numerous proteins that otherwise promote the ability of cancer cells to detach and invade. This includes EMT-regulating transcription factors as well as members of the regulatory network that allows dynamic reorganisation of the actin cytoskeleton. We have also found that miR-200 controls expression of an RNA-binding protein, QKI, which is responsible for regulating many alternative splicing events during EMT, including multiple components of the actin cytoskeleton-regulating network, and numerous circular RNAs.

27

Coordination of RNA Processing events in vivo

Karla Neugebauer[1]

1. Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA

Our goal is to understand how gene expression is controlled in living cells, using a combination of fluorescence imaging, biochemistry and molecular biology, as well as custom-designed RNA-Seq methods to observe dynamic nuclear functions *in vivo*. How is gene regulation achieved through pre-mRNA splicing, transcription, chromatin and the 3D organization of the cell nucleus? The use of budding yeast, mammalian tissue culture cells and zebrafish embryos as model systems has allowed us to investigate how transcription and splicing are coordinated. We have recently shown that the spliceosome completes exon-exon ligation as the 3' splice site emerges from Pol II. The proximity of the spliceosome and Pol II suggests physical and/or temporal cross-regulation among these machineries. Indeed, splicing feeds back to transcription by affecting elongation rates and or pausing; splicing also feeds forward to mRNA export by guiding mRNA export factors to mRNA binding. I will present unpublished data using long read sequencing of full length nascent RNAs in *Schizosaccharomyces pombe* showing the coordinated removal of introns within single transcripts as well as a dependency of polyadenylation cleavage on splicing. The findings that nuclear Cajal bodies (CBs) are sites of efficient spliceosomal snRNP assembly and essential for zebrafish embryogenesis are reflective of the propensity of the gene expression machinery to segregated – possibly through liquid-liquid phase separation – within the cell nucleus. I will discuss several models of nuclear function that could lead to the coordination of gene expression, which are suggested by the combination of these insights.

28

Enhancing the spatiotemporal sub-nuclear sequestration of RNA binding proteins by manipulating paraspeckle size

Archa Fox[1, 2], **Alina Naveed**[3, 2], **Jack Cooper**[1], **Ruohan Li**[1], **Steve Wilton**[4], **Sue Fletcher**[4], **Charlie Bond**[2], **Tao Liu**[5]

1. School of Human Sciences, University of Western Australia, Crawley, WA, Australia

2. School of Molecular Sciences, University of Western Australia, Crawley, WA, Australia

3. School of Human Sciences and Molecular Sciences, University of WA, Perth, WA, Australia

4. Centre for Comparative Genomics, Murdoch University, Perth, WA, Australia

5. Children's Cancer Institute, Kensington , NSW, Australia

Proteostasis of many important RNA binding proteins is critical for a variety of diseases including neurodegeneration and cancer. Given that the overall levels of these proteins can remain relatively constant, cells have developed other ways of regulating their function, including by spatiotemporal control of protein location. Paraspeckles are sub-nuclear RNA-protein bodies that form in response to cellular stress. Paraspeckle formation is triggered by enhanced transcription of a 23kb long noncoding RNA, NEAT1 (Nuclear Paraspeckle Assembly Transcript 1) that acts as a scaffold for binding by many different nuclear RNA binding proteins. This network of proteins then recruits additional proteins and RNAs to form a micron-scale subnuclear body, the paraspeckle.

We and others have shown that one function of paraspeckles is to 'sponge' up RNA binding proteins, effecting spatiotemporal sequestration of these proteins and thereby having a flow-on effect on their downstream targets. Here I will present new results on an antisense-oligonucleotide based method to enhance paraspeckle formation as a means of manipulating the sub-nuclear localisation of RNA binding proteins. We have specifically explored using this tool to manipulate paraspeckles in the context of the cancer neuroblastoma, as a means of sponging the oncogenic paraspeckle protein, NONO.

35

Integrative analysis of *in vivo* models of pancreatic cancer reveals complex mechanisms behind treatment failure and provides new tools for effective targeting

Marina Pajic[1]

1. Garvan Institute of Medical Research, Darlinghurst, NSW, Australia

Pancreatic cancer remains a highly lethal cancer where response is limited by both intrinsic and acquired chemoresistance. Understanding resistance mechanisms may therefore lead to improved therapeutic strategies. We have recently defined specific molecular subgroups of pancreatic cancer associated with pre-clinical and clinical response to select tailored treatment strategies [1-3]. Using robust patient-derived xenografts (PDXs) of pancreatic cancer, here we generated novel *in vivo* models for the study of intrinsic and acquired chemoresistance mechanisms to clinically-used agents, gemcitabine, mitomycin C, and cisplatin.

Whole genome sequencing and microarray profiling of gemcitabine-resistant tumours revealed complex but potentially targetable resistance mechanisms, including increased DNA repair through activation of PARP1, MCM genes and RRM1, and changes within the tumour microenvironment. Importantly, acquired resistance to gemcitabine was effectively reversed by a novel PARP inhibitor, rucaparib, indicating that combination therapy involving this low toxicity agent may be useful in treating gemcitabine-resistant tumours defined by high genomic instability. Similarly, modulation of key components of the tumour microenvironment with fasudil, as recently achieved [2], provided another effective way of reversing gemcitabine resistance.

Significance: Our findings demonstrate the promise of patient-derived xenograft models for the study of *in vivo* mechanisms of chemotherapy resistance and efficacy testing of novel agents for the treatment of human pancreatic cancer.

[1] Waddell N et al Nature (2015) 518(7540):495

[2] Vennin C et al Science Translational Medicine (2017) pii: eaai8504

[3] Chou A et al Gut (2017) pii: gutjnl-2017-315144 [epub ahead of print]

36

Spinraza: The first approved treatment for spinal muscular atrophy

Adrian Krainer[1]

1. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, United States

Nusinersen (Spinraza®), the first approved drug for spinal muscular atrophy (SMA), exemplifies a successful path from basic studies of pre-mRNA splicing mechanisms to an effective treatment for a devastating disease. This successful clinical application comes >120 years after the first description of the disease, and ~40 years after the discovery of RNA splicing and the first reported use of antisense technology.

SMA is a motor-neuron disease, caused by mutations in *SMN1*. Patients retain one or more copies of the nearly identical *SMN2* gene, which mainly expresses mRNA lacking exon 7, coding for an unstable protein isoform. The small amount of full-length mRNA and protein expressed from *SMN2* only partially compensates for the loss of *SMN1*. Together with Ionis Pharmaceuticals, we developed nusinersen, a splice-switching antisense oligonucleotide (ASO) that efficiently promotes *SMN2* exon 7 inclusion and restores SMN protein levels. Nusinersen hybridizes to intron 7 of the *SMN2* pre-mRNA, preventing binding of the splicing repressors hnRNPA1/A2 to a bipartite intronic splicing silencer; this in turn facilitates binding of U1 snRNP to the intron 7

5' splice site, resulting in enhanced exon 7 inclusion. Clinical trials of nusinersen in SMA patients, sponsored by Ionis and Biogen, began at the end of 2011. Based on the results of two phase-3 trials in infants with the most severe form of SMA, and in children with an intermediate form of SMA, respectively, Spinraza was approved by the FDA in December 2016, for all SMA types. It was also approved in Europe, Japan, and Australia in 2017.

We are continuing to explore aspects of SMA pathogenesis and treatment, using ASO therapy in SMA mouse models. We are also exploring prenatal ASO treatment, as it is likely that early intervention will maximize the clinical benefit.

37

Therapeutic alternative splicing: Making sense where there was none

Steve Wilton[1, 2], Sue Fletcher[1, 2]

1. Perron Institute for Neurological and Translational Science, Perth, WA, Australia

2. Centre for Comparative Genomics, Murdoch University, Murdoch, WA, Australia

Duchenne muscular dystrophy (MD), a common and serious form of childhood muscle wasting, typically arises from genomic deletions of one or more exons that disrupt the dystrophin reading-frame. Gene deletions that maintain the dystrophin reading-frame generally result in the milder Becker MD. Genotype:phenotype correlations clearly indicate that more than half the 79 dystrophin exons can be omitted from the mRNA without seriously compromising protein function. We have developed a splice intervention therapy using antisense oligonucleotides to specifically redirect dystrophin pre-mRNA processing. Targeted exon skipping can either re-frame a dystrophin mRNA corrupted by the loss of a frame-shifting exon or remove an exon containing a protein-truncating mutation.

Exondys 51, a morpholino oligomer designed to excise human dystrophin exon 51 from the mature mRNA, was granted accelerated approval by the US Food and Drug Administration. Relevant to about 10% of DMD mutations, *Exondys 51* is the first dystrophin restoring drug, the first exon-skipping drug, and the first morpholino oligomer to be approved for clinical use.

Designing clinical trials to evaluate drugs for rare diseases is challenging, especially when patient numbers are small and disease progression is relatively slow. There was vigorous and heated discussion within and outside the FDA after the accelerated approval of *Exondys 51*, and these concerns will be discussed and addressed. Therapeutic alternative splicing strategies must be tailored to different dystrophin mutations and Phase 3 clinical trials have been initiated to evaluate skipping of two other dystrophin exons.

Since most human genes undergo some form of splicing during expression, therapeutic alternative splicing could be relevant to many other inherited and acquired conditions, especially if mutated exons were known to be dispensable. We are currently exploring splice intervention therapies for spinal muscular atrophy, cystic fibrosis. Marfan's disease, Ehler-Danlos syndrome, adult-onset Pompe's disease, Epidermolysis bullosa, amyotrophic lateral sclerosis and multiple sclerosis.

38

Unavailable at time of print

Marylyn Ritchie[1]

1. University of Pennsylvania, Philadelphia, PA, United States

Unavailable at time of print

39

Developing iPSCs as a model for primate evolutionary genomics

Irene Gallego Romero[1]

1. University of Melbourne, Parkville, VIC, Australia

Functional genomics studies in primates have been consistently hindered due to the logistic and ethical limitations of obtaining samples – especially in the case of the critically endangered great apes, humanity's closest living relatives. But if we wish to truly understand the molecular background of human-unique traits, it is essential that we address these constraints. Reprogramming of somatic cell lines into induced pluripotent cell lines (iPSCs) from non-human primates, and their subsequent differentiation into otherwise unobservable tissue types, provides a transformative and highly versatile model system with which to examine the

contributions recent gene regulatory change has made to our species. We have previously generated a panel of chimpanzee (*Pan troglodytes*) iPSCs and thoroughly characterised their transcriptional and regulatory landscape, as well as the effects of reprogramming on gene expression. We find that expression and regulatory variation in iPSCs is consistently constrained relative to the somatic tissues we reprogrammed the cells from, and inter-species differences attenuated in magnitude, suggesting this state is highly conserved between the two species. I will also present preliminary data from targeted differentiation experiments which suggest that iPSCs are a viable method for identifying functional interspecies differences – and the molecular mechanisms that underlie them – but that on occasion a combination of batch and donor effects can overpower relevant trends. Taken together our findings illustrate both the promise and limitations of using this iPSC-modelling paradigm for interspecies comparisons.

40

The genome-wide rate and spectrum of somatic mutation in individual plants

Robert Lanfear[1], Adam J Orr[2], David Kainer[1], Alejandro Morales-Suarez[3], Amanda Padovan[1], Reed Cartwright[2]

1. Australian National University, Canberra, ACT, Australia

2. Arizona State University , Phoenix

3. Macquarie University, Sydney

Somatic mutations are thought to be fundamentally important to the ecology and evolution of plants, but we know very little about them because they are rare and difficult to detect. We have developed methods to reliably measure the genome-wide accumulation of somatic mutations in individual plants. By applying them to individual eucalypts of large stature, we are able to measure both the rate and spectrum of new mutations accumulating within individuals. In addition, we can pinpoint mutations of potential interest in understanding herbivore-resistant phenotypic mosaics. Our results have some interesting implications for plant ecology, evolution, and development.

41

Using scRNA-seq to construct developmental lineage relationships in the mouse mammary gland

Gordon K Smyth[1], Yunshun Chen[1], Bhupinder Pal[1], Jane E Visvader[1]

1. Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia

My lab has worked on breast cancer research for more than 12 years in collaboration with the lab of with Professors Visvader and Lindeman. In particular we have worked to understand the development of mammary cells and to identify the likely cells of origin for different types of breast cancer. In a recent study, we conducted scRNA-seq profiling of 20,000 mammary epithelial cells from mice at various life stages from pre-puberty to puberty, adulthood, pregnancy, estrus and diestrus. I will present some of the results form this study and will discuss some of the computational and statistical approaches used to interpret the data.

42

Whole Genome Sequencing for the Detection of Pathogenic Mitochondrial Mutations

Clare Puttick[1], Kishore R Kumar[1, 2], Ryan Davis[1, 2], Marcel E Dinger[1, 3], Carolyn M Sue[2], Mark J Cowley[1, 3, 4]

1. Kinghorn Centre for Clinical Genomics, Garvan Institute, Darlinghurst, NSW, Australia

2. Department of Neurogenetics, Kolling Institute of Medical Research, Royal North Shore Hospital and University of Sydney, St Leonards, NSW, Australia

3. St Vincent's Clinical School, University of New South Wales, Darlinghurst, NSW, Australia

4. Children's Cancer Institute, Randwick, NSW, Australia

Mitochondrial diseases (MD) represent one of the most common groups of inherited metabolic disorders, however due to their wide-ranging phenotypic spectrum they are notoriously difficult to diagnose. There are hundreds of mitochondria per cell, and pathogenic mutations can be present in 0-100% of their genomes, known as heteroplasmy. We performed Whole Genome Sequencing (WGS) on 250 patients with known or suspected MD. The aim was to evaluate the ability of WGS to detect mitochondrial mutations in blood. This could potentially avoid invasive muscle biopsies and simplify the diagnostic paradigm for MD.

In a 30-45x WGS genome we observed 4000-8000x coverage in the mitochondrial genome. We developed a dedicated bioinformatic pipeline, *mity*, to comprehensively identify SNVs and INDELs down to 0.3% heteroplasmy. *mity* identified ~39 mitochondrial variants per patient with a heteroplasmy of >1%, 30% of which

had a heteroplasmy <5%. We identified 71/79 known pathogenic mitochondrial variants identified prior to WGS, three of which had heteroplasmy < 1%. Pyrosequencing confirmed the remaining eight variants were absent in blood. Of the 67 m.3243A>G variants identified in blood, the heteroplasmy identified by *mity* and pyrosequencing was highly correlated.

Interestingly, there was no relationship between the Nijmegen Clinical Criteria Score for MD and pathogenic variant heteroplasmy, or the likelihood of a WGS diagnoses. In fact, 9 patients scoring 'not fulfilling' were diagnosed with pathogenic mutations with heteroplasmy up to 100%. We identified a number of individuals with complex clinical presentation, with more than one pathogenic mutation, which would have been missed using a more targeted approach. Finally, we diagnosed 4 previously undiagnosed patients with a pathogenic mitochondrial mutation.

mity's detection of low heteroplasmy mitochondrial variants, combined with an assessment of the nuclear genome, resulted in an overall diagnostic rate of 56%, further supporting the utility of WGS as a diagnostic test for MD.

43

Interpreting transcriptional heterogeneity in pluripotent stem cells using concerted computational approaches

Katie Willis[1], Thalia Chan[2], Jake Hobby-Royle[2], Petra Hajkova[3], Michael Stumpf[2], Kirsten R McEwen[2]

1. Imperial College London, Silwood Park, London, United Kingdom

2. Department of Life Sciences, Imperial College London, South Kensington, London, United Kingdom

3. MRC London Institute of Medical Sciences, London, United Kingdom

Transcriptional heterogeneity is a fundamental biological process impacting health and disease, yet the regulatory mechanisms behind this are poorly understood. Here we identify mechanistic processes of heterogeneity in pluripotent stem cells by manipulating the growth environment. We have developed new computational methods for analysing transcriptional heterogeneity and apply these to inform mechanisms of gene regulation.

First, we show how bulk population transcriptomics can be used to inform mechanisms of single-cell heterogeneity. Second, we have developed a method to infer rates of bursty gene transcription from single-cell RNA-seq transcriptomics. Third, we demonstrate the improved power of single-cell data for inference of biological network structure and have developed a new method to incorporate multiple layers of information, such as ChIP-seq data, into gene regulatory networks.

The complexity of single-cell heterogeneity can be unravelled using concerted computational approaches. Such methods are critical to understand the mechanistic processes regulating transcriptional heterogeneity and how these are disrupted in disease.

44

Recurrent regions of copy number variations in autism spectrum disorder are enriched for brain enriched coding and non-coding RNAs

Hamid Alinejad Rokny[1], Julian Heng[2], Alistair Forrest[1, 3]

1. Harry Perkins Institute of Medical Research, The University of Western Australia, Perth, WA, Australia

2. Curtin Health Innovation Research Institute, , Curtin University, Perth, WA, Australia

3. Division of Genomic Technologies, RIKEN Center for Life Science Technologies, Yokohama, Japan

Autism spectrum disorder (ASD) is a neurodevelopmental disorder with substantial phenotypic and etiological heterogeneity. 10-20% of cases are thought to be due to copy number variations (CNVs). To further identify the causal genes within these CNVs we developed a novel statistical framework to identify genomic regions that are recurrently deleted or duplicated in ASD. Analysis of CNV data from 19,663 autistic patients and 6,479 normal controls, identified 57 recurrently deleted and 62 recurrently duplicated regions associated with ASD, some of which show gender specific biases. Examining the phenotypes of patients with these 119 deletions and duplications in an independent cohort of patients with CNVs (DECIPHER) confirmed these regions are associated with ASD phenotypes, however also revealed distinct differences in the range of observed phenotypes for different duplications and deletions. To identify likely candidate genes within the significant genomic regions, we used the FANTOM5 expression atlas to prioritize those with enriched expression in brain. Of 830 coding genes found in the recurrently deleted regions and 518 in the duplicated regions, the expression of 37.3% and 37.4% are significantly enriched in brain. This is a significantly higher proportion than the 28.2% of all genes in the genome that have brain enriched expression. Gene ontology analyses also identified distinct differences in the biological process annotations of brain enriched genes within deleted and duplicated regions. Lastly we report for the first time 510 and 296 long non-coding RNAs found in deleted and duplicated regions

respectively which have enriched expression in brain. Interestingly nine of our significant regions have only significant brain enriched lncRNA suggesting that for these regions the lncRNA is the most likely candidate for the observed phenotype. Our analyses highlight the diversity of genetic lesions that contribute to ASD and provide new genetic evidence for sub classification within the spectrum of ASD.

45

TAD cliques shape the 4-dimensional human genome during lineage-specific differentiation

Philippe Collas[1], Jonas Paulsen[1], Maxime Nekrasov[2], David Tremethick[2]

1. University of Oslo, Oslo, Norway

2. Australian National University, Canberra

Dynamic alterations in genome conformation establish blueprints of developmental gene expression. These changes reflect a 4-dimensional (4D) orchestration of temporal gene regulation in the nucleus space. We report large-scale 4D changes in genome conformation occurring during commitment of human mesenchymal stem cells into adipogenic and neuronal lineages. We combine maps of topologically-associated domains (TADs) and lamin-associated domains (LADs), together with a 3D genome modeling tool (Chrom3D) [1], to infer TAD position in space during bi-lineage differentiation. Whereas TAD boundaries are stable along the linear genome, TADs show temporal dynamics in space. We unveil the formation of multiple TAD-TAD associations into multi-megabase repressive hubs, or TAD cliques. Most TAD cliques are constitutive; however others assemble or break apart in a lineage-specific manner. We highlight adipogenic- and neurogenic-specific transient TAD cliques engulfing or releasing cell cycle, adipogenic or neurogenic loci. Through associations with the nuclear lamina, TAD clique assembly or growth draws genic regions into a repressive environment at the nuclear periphery. LADs appear as facilitators of long-range TAD repositioning in the nucleus space that establish new TAD clique patterns. Lineage-specific dynamic TAD cliques highlight a new level of developmental genome organization. Our findings provide an overview of large-scale changes in the 4D nucleome during lineage-specific differentiation. Lastly, we are developing virtual reality (VR) genomics, enabling interactive spatial exploration of the genome. VR genomics shows potential for analysis and educational purposes.

[1] Paulsen et al. 2017. Genome Biol 18, 21.

46

Loss of Suv39h1 and h2 histone methyltransferases results in disruption of chromatin organisation which leads to immune progeria

Christine R Keenan[1, 2], Nadia Iannarella[1], Hannah D Coughlan[1, 2], Alexandra L Garnham[1, 2], Gordon K Smyth[1, 3], Rhys S Allan[1, 2]

1. Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia

2. Department of Medical Biology, The University of Melbourne, Parkville, VIC, Australia

3. School of Mathematics and Statistics, The University of Melbourne, Parkville, VIC, Australia

Epigenetic changes are considered a hallmark of aging. However, the molecular drivers for these changes are not well described. Suv39h1 and Suv39h2 enzymes are responsible for trimethylating lysine 9 of histone 3 (H3K9me3) leading to stable gene silencing through the formation of heterochromatin via recruitment of heterochromatin 1 proteins. Suv39h1 expression decreases with age, and Suv39h2, whilst ubiquitously expressed during embryogenesis, is expressed at low levels in all adult tissues except the testes.

Deletion of Suv39h repressive enzymes would be expected to cause aberrant gene activation. However, remarkably, RNASeq on double-positive thymocytes, precursors of both $CD4^+$ and $CD8^+$ T lymphocytes, from mice deficient in both Suv39h enzymes (Suv39dn) showed widespread gene repression. These Suv39dn cells also show disordered heterochromatin and a marked loss of LaminB1 protein leading to altered nuclear morphology. Interestingly, in contrast to the majority of differentially-expressed genes, LaminB1-chromatin immunoprecipitation (ChIP) of wildtype thymocytes reveals that the majority of lamina-interacting genes are significantly up-regulated suggesting that disruption of nuclear lamina-tethering leads to selective gene activation for this subset of genes.

Together with molecular changes, aging leads to well-characterised changes in adaptive immunity resulting in reduced immune function and an increased frequency of infections among older individuals. These changes include reduced haematopoietic stem cell (HSC) potential, reduced $CD8^+$ lymphocyte numbers, and an increased frequency of spontaneous antigen-naïve memory cells. Importantly, these age-associated immune features are recapitulated in Suv39dn where we see reduced HSC potential revealed through poor reconstitution of bone-marrow chimeras by Suv39dn cells. Furthermore, reduced $CD8^+$ T lymphocyte numbers and increased memory $CD8^+$ T cells were observed.

Taken together, these results suggest that decreased levels of Suv39h1 and h2 in aging drives a loss of heterochromatin and nuclear lamina integrity which in turn drives the loss of immune function and susceptibility of older individuals to infections.

47

No kissing in the nucleus: Genome-wide analysis reveals no evidence of trans chromosomal regulation of mammalian immune development

Timothy Johanson[1, 2], Hannah Coughlan[1, 2], Aaron Lun[1, 2], Naiara Bediaga[1, 2], Gaetano Naselli[1, 2], Alexandra Garnham[1, 2], Leonard Harrison[1, 2], Gordon Smyth[1, 3], Rhys Allan[1, 2]

1. The Walter and Eliza Hall Institute, Parkville, Vic, Australia

2. Dept. of Medical Biology, The University of Melbourne, Melbourne, Vic, Australia

3. Department of Mathematics and Statistics, The University of Melbourne, Parkville, Victoria, Australia

It has been proposed that interactions between mammalian chromosomes, or transchromosomal interactions (also known as kissing chromosomes), regulate gene expression and cell fate determination. We aimed to identify novel transchromosomal interactions in immune cells by high-resolution genome-wide chromosome conformation capture. Although we readily identified stable interactions in *cis,* and also between centromeres and telomeres on different chromosomes, surprisingly we identified no gene regulatory transchromosomal interactions in either mouse or human cells, including previously described interactions. We suggest that advances in the chromosome conformation capture technique and the unbiased nature of this approach allowed more reliable capture of interactions between chromosomes than previous methods. Overall our findings suggest that, contrary to the dogma, stable transchromosomal interactions that regulate gene expression are not present in mammalian immune cells and that lineage identity is governed by *cis*, not *trans* chromosomal interactions.

48

Inhibition of a K9/K36 Demethylase by an H3.3 Point Mutation Found in Paediatric Glioblastoma

Hsiao Voon[1], Maheshi Udugama[1], Wendi Lin[1], Linda Hii[1], Ruby Law[1], Partha Pratim Das[1], Jeff Mann[1], Lee Wong[1]

1. Monash University, Clayton, VIC, Australia

A number of oncogenic histone point mutations have been identified across a host of cancer studies. We report that the H3.3 G34R (glycine to arginine) substitution mutation found in paediatric gliomas, causes widespread changes in H3K9me3 and H3K36me3 by interfering with the Kdm4 family of K9/K36 demethylases. Expression of a targeted single-copy of H3.3 G34R was sufficient to induce chromatin alterations which genocopied a Kdm4 a/b/c triple-knockout. *In vitro* and *in vivo* immunoprecipitation assays demonstrated that the H3.3 G34R mutant binds Kdm4 with high affinity while simultaneously inhibiting enzymatic activity. These studies show that H3.3 G34R acts in a dominant negative fashion by inhibiting Kdm4, triggering genome-wide alterations to promote oncogenesis. We propose a general model where oncogenic histone mutations alter the genome through interactions with epigenetic erasers, writers and readers.

49

Epigenetic principles of reprogramming in vivo

Petra Hajkova[1]

1. MRC London Institute of Medical Sciences, Imperial College London, London, UK

During mouse embryonic development early postfertilisation zygotes and the developing primordial germ cells (PGCs, the precursors of gametes) undergo global epigenetic reprogramming. This process involves genome-wide erasure of DNA methylation as well as global changes in chromatin structure and histone modifications. Despite the efforts of numerous research teams the molecular mechanisms underlying these developmental reprogramming processes remain elusive.

I will present our recent results regarding the dynamics of DNA modifications and chromatin during the epigenetic reprogramming in mouse gonadal primordial germ cells (PGCs). I will also discuss our current understanding of mechanistic links between the global DNA demethylation and the execution of the germline developmental programme.

50

The role of co-activator complexes in regulating RNA polymerase II transcription

Laszlo Tora[1]

1. Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC/CNRS), Illkirch, ALSACE, France

Initiation of transcription by RNA polymerase II (Pol II) is the outcome of a number of sequential events beginning with the binding of activators to their binding sites, which will trigger the recruitment of coactivator complexes and general transcription factors (GTFs) at promoters to allow the loading of Pol II into the preinitiation complex. In this process, coactivators play multiple crucial roles. SAGA (Spt-Ada-Gcn5-Acetyltransferase) and the ATAC (Ada-Two-A-Containing) complexes are two functionally distinct, but related, coactivator complexes with several enzymatic activities. The inactivation of yeast SAGA induced a strong decrease of nascent transcription from all expressed genes demonstrating a general requirement of SAGA at all transcribed yeast genes[1].

To better characterize the behavior of GTFs, Pol II, SAGA and ATAC in living human cells we analyzed the dynamic nature of GFP-tagged factors with fluorescence recovery after photobleaching (FRAP), fluorescence loss in photobleaching (FLIP) and fluorescence correlation spectroscopy (FCS). FRAP and FLIP measurements indicated that TFIIB, TFIID, ATAC and SAGA subunits are highly dynamic and exhibit only transient interactions with the chromatin with no detectable immobile fractions. FCS measurements indicated that TFIIB, TFIID, ATAC and SAGA have two distinct diffusing populations in the nucleus: “fast” (free complexes) and a “slow” population (chromatin interacting complexes). Inhibition of transcription and reduced levels of histone H3K4me3 decreased the “slow”, chromatin interacting, population of SAGA and ATAC, demonstrating that the equilibrium between free and chromatin interacting SAGA and ATAC complexes is regulated by the active transcription-dependent chromatin landscape[2].

In addition, a novel method will be presented allowing measurements of dynamic changes in the distribution of endogenous transcription factors (Pol II), and specific modified histones (gH2AX), at the nanometre scale[4].

1. 1Baptista et al. Mol. Cell (2017).
2. 2Vosnakis et al. EMBO J. (2017).
3. 3Conic et al, JCB, in press. (2018)

51

The remote control of gene expression

Wendy Bickmore[1]

1. MRC Human Genetics Unit, University of Edinburgh, Edinburgh, MIDLOTHIAN, United Kingdom

It is now appreciated that long-range enhancers - found as far away as 1 megabase from their target gene and located either in intergenic regions or in introns - are key in controlling the precise spatial and temporal expression of genes. In contrast to the 20,000 or so genes in our genome, there may be hundreds of thousands, or even millions, of enhancers. Deletion, translocation or point mutations can abrogate the function of these elements in Mendelian diseases associated with severe phenotypes. However, the majority of human genetic variation associated with common and complex disease and quantitative traits also maps to intergenic regions that are likely the site of enhancers. Therefore, lessons learnt from studying enhancer dysfunction in rare disease will be important for an understanding of milder phenotypes.

It is hard to envisage how distant enhancers function if one only considers the genome as a linear DNA sequence. Rather, three-dimensional chromatin folding in the nucleus must play a fundamental role in enhancer-promoter communication. I will describe our work using different experimental approaches to investigate and manipulate the three-dimensional folding of the mammalian genome at genetically defined long-range regulatory elements. Models of enhancer function from chromatin looping to linking and tracking will be discussed.

52

The expanding clinical utility of cancer genomes

Kenneth W Kinzler[1]

1. John Hopkins Medicine, Baltimore, MD, United States

The last three decades has seen a revolution in our ability to characterize cancer genomes and apply this knowledge to clinical questions. In oncology, these applications began with inherited predispositions to cancer and progressed to applications of somatic cancer genomes. Recent technological advances now allow the

sensitive detection of trace levels of altered genomes and introduced the era of mutations as biomarkers (e.g. liquid biopsies). This lecture will briefly review the progression of clinical applications and then focus on the utility of trace genomes for the earlier detection of cancer.

101

Investigating the function of 5-methylcytosine in mRNA biology

Ulrike Schumann[1], Chikako Ragan[1], Tennille Sibbritt[1, 2], Andrew Shafik[1, 3], Maurits Evers[1], Thomas Preiss[1]

1. John Curtin School of Medical Research, ANU, Canberra/Acton, ACT, Australia

2. Children's Medical Research Institute, Sydney Medical School, University of Sydney, Sydney, NSW

3. Boston University School of Medicine, Department of Biochemistry, Boston, Massachusetts, USA

Here we present a transcriptome-wide analysis of the role of 5-methylcytosine in RNA biology, opening a gateway into fully unraveling the role of this mark in cellular processes.

5-methylcytosine is a post-transcriptional modification common in tRNA and rRNA, playing an important role in tRNA function and stability as well as ribosome biogenesis and assembly. The 5-methylcytosine modifications are primarily mediated by the NOP2/Sun (NSUN) family of RNA methyltransferases, which consists of seven members. NSUN2, the best understood member, targets tRNAs for methylation, which plays an important role in decoding, tRNA stability and stress response. NSUN2 has also been shown to target other RNA biotypes such as snoRNAs, lncRNAs and mRNAs, but the role of 5-methylcytosine in these remains poorly characterised.

Focusing specifically on mRNAs, 5-methylcytosine could affect processes such as splicing, stability, RNA degradation or miRNA targeting as well as translation. Recently roles for 5-methylcytosine in mRNA translation and stability was proposed for specific mRNAs. To gain an understanding of mRNA regulation by 5-methylcytosine at a transcriptome-wide level, we analyzed the effect of loss of 5-methylcytosine, by means of NSUN2 knock-down, on mRNA steady-state level. A large proportion of mRNAs affected by loss of NSUN2, showing either up- or down-regulation, are predicted to also harbour 5-methylcytosine. Further, we investigated the extent of ribosome association of modified mRNAs compared to their unmodified counterparts, by employing bsRNA-seq of RNA isolated from lowly to highly translated regions of a polysome gradient. Utilizing clustering analyses we reveal that a subset of mRNAs show either positive or negative correlation of translation state with 5-methylcytosine modification at specific sites.

We discuss our overall findings and highlight the regulation of specific examples in detail.

102

BRG1 maintains chromatin at active enhancers and the expression of proliferation genes in prostate cancer

Katherine A Giles[1], Cathryn M Gould[1], Joanna Achinger-Kawecka[1, 2], Phuc-Loi Luu[1], Shalima S Nair[1], Susan J Clark[1, 2], Phillippa C Taberlay[1, 2, 3]

1. The Garvan Institute, Darlinghurst, NSW, Australia

2. St Vincent's Clinical School, UNSW Sydney, Sydney, NSW, Australia

3. School of Medicine, Faculty of Health, University of Tasmania, Hobart, TAS, Australia

ATP-dependent chromatin remodelling is required for all aspects of chromatin biology including transcription, DNA repair, DNA replication, and cell cycle progression. Deregulation of *SMARCA4* (BRG1), an essential catalytic subunit of the SWI/SNF chromatin-remodelling complex, contributes to a wide range of malignancies. Furthermore, clinical studies have revealed that BRG1 is over expressed in prostate cancer and positively correlated with disease progression and invasiveness. However, the mechanisms underlying the involvement of BRG1 are unknown. Here we investigated the consequences of BRG1-depletion on the epigenome in LNCaP prostate cancer cells. BRG1 is over expressed in LNCAP cells compared to normal prostate epithelial cells (PrEC) and therefore provides an ideal model for molecular investigation. We found that BRG1 (ChIPseq) binding is enriched at active DNA regulatory elements including enhancers and promoters. Depletion of BRG1 causes a rapid compaction of chromatin (ATACseq) at active enhancers and at Polycomb marked regions, but not at promoters. This occurs concomitant with a reduction in H3K27ac and H3K4me1 (ChIPseq) leading to a significant reduction in the number of active enhancers. Despite the substantial effect on the enhancer landscape, we detected only a modest effect on gene expression (RNAseq), suggesting that vigorous compensatory mechanisms exist. The majority of gene expression changes were down regulated genes enriched for GO terms including cell cycle, DNA replication and DNA metabolism. Complementary FACS cell cycle analysis and determined that BRG1 loss reduces the number of cells in S-phase and increases cells in G1, corresponding with a reduction in proliferation. Taken together we find that BRG1 has a crucial role in

maintaining open chromatin at active enhancers, which likely underpins the atypical gene expression signature and altered growth capacity of prostate cancer cells.

103

Tools for preprocessing and benchmarking single cell RNA-sequencing data

Luyi Tian[1], Shian Su[1], Daniela Amann-Zalcenstein[1], Christine Biben[1], Shalin Naik[1], Matthew E Ritchie[1]

1. Walter Eliza Hall Institute of Medical Research, Parkville, VIC, Australia

Single cell RNA-sequencing (scRNA-seq) technology allows researchers to profile the transcriptomes of thousands of cells simultaneously. The high throughput nature of this approach leads to a more complex data structure, with a mix of designed and random barcodes used to identify different cells and molecules that must be efficiently dealt with to arrive at a count matrix for further analysis. For downstream analysis, there is already a large number of methods available for scRNA-seq data that deal with problems ranging from normalization to clustering and trajectory analysis. Efforts to comprehensively benchmark different methods are still in their infancy and are currently hampered by the lack of gold-standard data sets.

To address these issues we developed new software tools and benchmarking data sets. For data preprocessing, the *scPipe* R/Bioconductor package was created to take raw sequence reads from FASTQ files generated by different protocols (including CEL-seq, MARS-seq, Chromium 10X and Drop-seq) and arrive at a gene count matrix for downstream analysis. *scPipe* performs demultiplexing, UMI deduplication, alignment and gene counting. It also aids in quality control by generating plots of a number of key quality metrics and robust outlier detection to remove poor quality cells.

To improve our ability to benchmark analysis methods, we designed and generated a number of scRNA-seq control data sets. Using a mixture design, cells and RNA from three lung adenocarcinoma cell lines were combined in different ratios to create known populations and pseudo-trajectories. Data was generated using CEL-seq and Chromium 10x protocols. A benchmarking software platform will be developed to facilitate methods comparisons using common diagnostics and make it easier to select optimal analysis methods for different tasks.

These software and data will be made freely available to help researchers process their raw data and guide the selection and development of better scRNA-seq analysis pipelines.

104

Sequential occupation of the nuclear periphery by H2A.Zac and H3K9me2 accompanies pluripotency loss in Human embryonic stem cells

Georgia R Kafer[1, 2], Peter M Carlton[2]

1. The Children's Medical Research Institute, Westmead, NSW, Australia

2. Graduate School of Biostudies, Kyoto University, Kyoto, Japan

The differentiation of a pluripotent cell into a cell of a specialised lineage is accompanied by extensive changes in the expression of pluripotency and lineage-specific genes. Gene expression changes can be mediated by both chromatin modifications which make DNA more or less accessible to transcription factors and the physical movement of gene loci into silent or active nuclear compartments. Although changes in chromatin architecture are critical for gene expression control, we currently do not fully understand the role of chromatin dynamics and nuclear subcompartments in facilitating differentiation. The bulk of existing work has been generated through genome-wide techniques which, while able to reveal information about chromatin changes at high genomic resolution, fails to show both changes in 3D nuclear architecture and any subtle cell-to-cell variation. In this study we have carefully analysed high-resolution images of hundreds of individual cells in an attempt to assay the nuclear patterning of differently modified histones during early phases of pluripotency loss and the beginning stages of terminal lineage commitment. We documented dynamic changes of modified histones in different nuclear compartments in hESCs from both primed and “naïve” pluripotent states in addition to hESCs which are differentiating into trophoblast like (TBL) cells via supplementation with BMP4 while simultaneously inhibiting FGF signaling and the Activin/Nodal pathway. We report that two histone modifications, the di-methylation of histone 3 at lysine 9 (H3K9me2) and the acetylation of the histone variant H2A.Z (H2A.Z.ac) show movement towards and away from the nuclear periphery as hESCs transition from naïve to primed and differentiated states. Interestingly, we show that the HDAC complexes are key controlling factors in the establishment of chromatin environments, especially at the nuclear periphery, and that deacetylation is critical to the establishment of chromatin environments in differentiated cells.

105

Intron retention redefines post-transcriptional gene regulation in vertebrate species

Ulf Schmitz[1, 2], Natalia Pinello[1, 2], William Ritchie[3], Maria Cristina Keightley[4], Shaniko Shini[5], Graham J. Lieschke[4], Justin J-L Wong[1, 2], John E.J. Rasko[1, 2]

1. Centenary Institute, Camperdown, NSW, Australia

2. Sydney Medical School, University of Sydney, Sydney, NSW

3. UMR 5203, CNRS, Montpellier, France

4. Australian Regenerative Medicine Institute, Monash University, Clayton, Victoria, Australia

5. Faculty of Medicine, The University of Queensland, St Lucia, Queensland, Australia

While intron retention (IR) is now widely accepted as an important mechanism of mammalian gene expression control, it remains the least studied form of alternative splicing. To delineate conserved features of IR, we performed an exhaustive phylogenetic analysis in a highly purified and functionally defined cell type comprising neutrophilic granulocytes from five vertebrate species spanning 430 million years of evolution.

Our RNAseq-based analysis suggests that IR increases gene regulatory complexity, which is indicated by a strong anti-correlation between the number of genes affected by IR and the number of protein-coding genes in the genome of individual species. Our results confirm that IR affects many orthologous or functionally related genes in granulocytes. Further analysis uncovers new and unanticipated conserved characteristics of intron-retaining transcripts. We find that intron-retaining genes are transcriptionally co-regulated from bi-directional promoters. Intron-retaining genes have significantly longer 3' UTR sequences, with a corresponding increase in microRNA binding sites, some of which include highly conserved sequence motifs. This suggests that intron-retaining genes are highly regulated post-transcriptionally.

Our study provides unique insights concerning the role of IR as a robust and evolutionary conserved mechanism of gene expression regulation. Our findings enhance our understanding of gene regulatory complexity by adding another contributor to evolutionary adaptation.

107

Identifying Hepatitis B virus integration events on Mouse genome

Waruni Abeysekera[1, 2], Zina Valaydon[1, 4, 3], Marc Pellegrini[1, 2], Gordon Smyth[1, 5]

1. Walter and Eliza Hall Institute of Medical Research, Melbourne, VICTORIA, Australia

2. Department of Medical Biology, The University of Melbourne, Parkville, VICTORIA, Australia

3. Department of Medicine, Eastern Hill Academic Centre , The University of Melbourne, Parkville, VICTORIA, Australia

4. Department of Gastroenterology, St Vincent's Hospital, Fitzroy, VICTORIA, Australia

5. School of Mathematics and Statistics, The University of Melbourne, Parkville, VICTORIA, Australia

Hepatitis B virus (HBV) causes the most common liver infection and is known as a leading risk factor for liver cirrhosis and hepatocellular carcinoma. More than 300 million people worldwide are estimated to be infected with this virus. HBV is a double stranded DNA virus of the family *Hepadnaviridae* that has the ability to integrate into host chromosomes when infected. The mechanism of the HBV DNA integration into host genome remains ambiguous, yet it is believed that this integration occurs during the virus replication. Here, we describe two approaches to identify HBV integration events in the mouse genome using RNA sequencing (RNA-Seq) data of HBV infected mouse liver samples. In the first approach, raw paired-end (PE) reads of RNA-Seq data are mapped to a hybrid reference genome built by concatenating mouse mm10 genome and HBV genome. A uniquely mapped of a PE read with one end mapped to mouse genome and the other end mapped to HBV genome is classified as a mouse-HBV chimeric PE read and is considered as an identified virus integration event. Such chimeric reads/integration events are annotated by Ref-Seq gene annotation and reads with same integration events are clustered together. Approach 2 consists of two step alignment of the RNA-Seq data, where the chimeric PE reads are identified by first mapping PE reads to the HBV genome, and the resulting mapped viral reads are then re-mapped to the mouse genome. These two approaches are illustrated by an RNA-Seq data of 12 mouse liver samples infected with 3 conditions of HBV, namely precore mutant, basal core promoter mutant and wild type.

108

Developing a novel embryonic lung progenitor culture system: technique and applications

Casey Ah-Cann[1, 2], Marie-Liesse Asselin-Labat[1], Matt Richie[1], Kelsey Breslin[1], Joy Liu[1], Tamara Mclennan[1], Charity Law[1], Marnie Blewitt[1]

1. Walter and Eliza Hall Institute, Parkville, VIC, Australia

2. Department of Medical Biology, The University of Melbourne, Melbourne, VIC, Australia

Development of the lung is a highly regulated and complex process that is not fully characterised. Lung development starts at embryonic day (E)9.5 in the mouse where progenitor cells drive morphogenesis of lung lobes and future airways. Progenitor cells express Sox9 but how they are regulated remains to be elucidated. We have developed a novel *in vitro* assay to expand Sox9-positive cells from E11.5 lung. These cells form pneumospheres and maintain their progenitor activity, enabling the study of factors regulating proliferation and differentiation. RNAseq analysis of pneumospheres overtime revealed an increase in Sox9 expression, and Sox9-downstream targets such as Clusterin and Annexin A4. Alveolar differentiation markers such as surfactant protein c and Rage are downregulated, indicating an enrichment for progenitor cells. Pneumospheres can be maintained in culture for several weeks enabling us to perform genetic or drug screens.

Using this platform, we performed a short-hairpin RNA (shRNA) knockdown screen, targeting 130 genes involved in enzymatic epigenetic regulation. Sequencing at three time-points enabled identification of hairpins that changed overtime. The top genes were validated using *in vitro* and *ex vivo* culture systems to determine their role in lung progenitor cells and branching morphogenesis.

These experiments identified Aurora kinase B (*Aurkb*) as an interesting candidate gene. Aurkb functions as a regulator of cell cycle and epigenetic control through phosphorylation of histones. Disruption of Aurkb by shRNA or by chemical inhibition abrogates growth of lung progenitor cells and causes defects in cell cycle. Future studies aim to decipher the role of Aurkb in lung development using mice carrying a conditional allele of Aurkb.

These studies in the embryo help us understand how developmental genes can become dysregulated later in life leading to disease. Our work highlights the power of interrogating lung development to provide insight into lung diseases and develop novel therapeutic targets.

109

RNA processing is critical for the generation of mature megakaryocytes capable of producing platelets

Tanveer Ahmad[1, 2], Shen Heazlewood[2], Brenda Willilams[2], Jess Hatwell-Humble[2], Monika Mohenska[1], Mirana Ramialison[3, 1], Belinda Guo[4], Susie Nilsson[3, 2], Minna-Liisa Anko[1]

1. Monash University, Clayton, VIC, Australia

2. Manufacturing, CSIRO, Melbourne, VIC, Australia

3. Australian Regenerative Medicine Institute, Melbourne, VIC, Australia

4. University of Western Australia, Perth, WA, Australia

Megakaryocytes (MK) are bone marrow cells responsible for platelet production. During maturation, MK undergo DNA replication without cytokinesis to become large in size (>65µm) and polyploid (≤128 N) but the molecular mechanisms underlying MK maturation are poorly understood. Our work has revealed that MK of individual ploidies exist as two distinct populations distinguished by set of cell surface markers such as CD41 and CD61. When CD41/CD61^{bright} MK that are larger in size and have a distinct complex/granular cytoplasm were transplanted into WT mice, they released 500x more platelets than CD41/CD61^{dim} MK. Intriguingly, we have identified a key role for the RNA binding protein Serine-arginine Rich Splicing Factor 3 (SRSF3) in megakaryopoiesis. In *Srsf3-Pf4-Cre* mice, where SRSF3 expression is ablated in MK, the CD41/CD61^{bright} MK population is largely missing, resulting in severe thrombocytopenia with a 90% reduction in platelets. This together with our ultrastructural analysis of *Srsf3*-null MK using transmission electron microscopy supports our hypothesis that CD41/CD61^{bright} MK subpopulation is responsible for platelet production. RNA-sequencing analysis of 8N versus >16N *Srsf3*-null and WT MK demonstrated that RNA processing mediated by SRSF3 plays a central role in tuning the MK transcriptome. We identified a large shift in RNA repertoire during MK maturation that did not take place in *Srsf3*-null MK. Moreover, platelets derived from *Srsf3-Pf4-Cre* mice were morphologically and functionally abnormal and had 30 times more RNA compared to WT platelets. Platelet RNA-sequencing showed that the *Srsf3*-null platelets had greatly abnormal RNA composition, with hundreds of up- and downregulated RNAs. Our work sheds light into the role of RNA processing in megakaryopoiesis and platelet biogenesis which may help in identifying new approaches to tackle thrombogenic and haematopoietic disorders.

110

Visualisation of telomeres by telomere fibre-FISH

Joshua AM Allen, Alexander P Sobinoff, Hilda A Pickett

Telomeres are nucleoprotein complexes that cap the ends of the linear eukaryotic chromosomes. Telomeres comprise repetitive non-coding DNA sequences (TTAGGG repeats in vertebrates) bound to the hexameric protein complex, shelterin. Telomeric DNA is eroded with each round of cell division, eventually resulting in cellular senescence. To counteract this process and achieve proliferative immortality, cancer cells must activate a telomere maintenance mechanism. The majority of cancers activate the ribonucleoprotein enzyme telomerase, while a smaller proportion engage the Alternative Lengthening of Telomeres (ALT) pathway. ALT is a homology-directed repair pathway that utilises telomeric templates for repeat extension, but the precise protein requirements of this complex multi-component pathway remain elusive

We have established telomere fibre-fluorescence *in situ* hybridisation (TFF) to visualise (i) the distribution of individual telomere lengths in a cell population, and (ii) the frequency and length of telomere extension events at individual telomeres. By employing a constant stretching factor of 2 kb/μm to stretch DNA fibres, followed by hybridisation to a telomere-specific PNA probe, we have measured telomere length in ALT cancer cells stably overexpressing the BLM helicase or the SLX4 structure-specific endonuclease. We identified telomere lengthening in BLM overexpressing ALT cell lines, and telomere shortening in SLX4 overexpressing ALT cell lines, and demonstrated the requirement for the BLM-TOP3A-RMI1-RMI2 (BTR) complex for ALT-mediated telomere synthesis. Our data are consistent with ALT being a conservative DNA replication process, analogous to break-induced replication. We are currently using TFF to investigate other protein components of the ALT pathway.

111

Characterisation of NHL-2 in small RNA pathways of *Caenorhabditis elegans.*

Joshua Anderson[1], Rhys Colson[1], Julie Claycomb, Peter Boag[1]

1. Monash University, Clayton, VIC, Australia

There is an increasing number of small RNA pathways that are intimately involved in the regulation of gene expression. These small RNA pathways are found to be critical for maintaining genome organisation and have also been shown to regulate the genome across generations through epigenetic mechanisms. In *C. elegans* a diverse repertoire of small RNA regulatory pathways operate in both the soma and germline that share similarities in biogenesis and mechanism of action, but diverge in targeting and regulatory role. The diverse functions of small RNA pathways include licensing of transcription and genome protection by limiting deleterious transcription or gene expression (genome protection from foreign elements). Aside from the core factors such as Argonaute's and their mediating proteins (ALG/AIN), there exist a series of modulating co-factors. Usually the presence of certain so-called co-factors, in addition to the core factors, is pathway dependent for regulating the efficiency of either biogenesis, targeting or action of these small RNAs. One such co-factor defined in the miRNA pathway is the TRIM-NHL protein, NHL-2, which was studied for roles in miRISC activity of *let-7* and *lsy-6* targets in the soma.

We have identified that NHL-2 holds more functions than just within miRISC activity, of most significance is a fundamental role in germline functions and reproductive capacity. We show a clear requirement for *nhl-2* in germline immortality and clear roles in transgenerational epigenetic inheritance. Further to the somatic and germline functions we show a role in the nuclear RNAi pathway, which makes *nhl-2* a promiscuous gene in opposing (licensing and repressive chromatin marks) RNA regulatory pathways. This is supported through biochemical studies showing physical interactions with key Argonaute proteins and additionally through small RNA-seq data. Collectively, we show a promiscuous modulating co-factor that forms a hub of gene regulatory activity in both the germline and soma.

112

Small RNA profiling of the reproductive tissues of the honeybee *Apis mellifera*

Alyson Ashe[1], Paul Yound[2], Gabriele Buchmann[1], Emily Remnant[1], Catherine Suter[2], Benjamin Oldroyd[1]

1. University of Sydney, Sydney, NSW, Australia

2. Victor Chang Cardiac Research Institute, Sydney

Small RNA molecules such as miRNAs, piRNAs and siRNAs have become recognised as vital in the regulation of many important processes such as the fine-tuning of gene expression during development, and repression

of transposons. Here we present the first in-depth characterisation of small RNAs present in the reproductive tissues of the honeybee. We have performed small RNA sequencing on ovaries, spermatheca, sperm, fertilised and unfertilised eggs and testes.

364 novel miRNAs were detected, with the majority coming from the testes and spermatheca samples. The expression pattern of known and novel miRNAs differs dramatically between tissue types with semen in particular having a highly divergent miRNA expression pattern.

piRNAs are a class of small RNA between 24-31 nt long with a strong 5'U bias and 2'O-methylation at the 3' end. They are generally void of sequence conservation and the ability to form hairpin structures from flanking genomic DNA, and thus differ greatly from miRNAs. They are usually transcribed from both large, repeat-rich, uni- and bi-directional clusters distributed throughout the genome. Their primary function differs based on tissue type, and can be either suppression of transposon activity or regulation of endogenous gene expression. Using custom scripts we have identified a total of 168 piRNA clusters. The clusters appear to be non-randomly distributed throughout the genome. The average cluster size is 10kb, and the average length of sequences located within the clusters is 28bp.

Finally, the semen samples are composed of approximately 60% tRNA fragments (tRFs). tRFs are a highly conserved, ancient type of small RNA molecule that have been suggested to have a role in the transmission of epigenetic signals between generations in mice. Surprisingly, the tRFs found in honeybee semen are almost identical to those implicated in epigenetic inheritance in mouse semen. The implications of this finding will be discussed.

113

The phosphorylation of H3.3 Serine 31 and its role in heterochromatin formation

Vinod Benjamin[1], Hsiao Voon[1], Maheshi Udugama[1], Lyn Chan[1], Linda Hii[1], Fiona Chang[1], Jeff Mann[1], Lee Wong[1]

1. Monash University, Clayton, VICTORIA, Australia

Histone H3.3, a variant of H3, is encoded by 2 genes: *H3F3A* and *H3F3B.* H3.3 has a Serine at residue 31 which is evolutionarily conserved and is known to be phosphorylated during metaphase. The enrichment of H3.3S31ph has been observed at heterochromatic repeats including the telomeres and pericentric satellite DNA. The presence of H3.3S31ph at these repeats is dependent on ATRX-mediated H3.3 loading. To date, little is known about the function and regulation of H3.3S31ph. In this study, we aim to define the function of H3.3S31ph, in particular, its role in telomeric chromatin assembly. We generated *Cre*-mediated H3.3 knockout mouse embryonic stem cell (ESC) lines expressing phospho-mimic (H3.3S31E) and phospho-null (H3.3S31A) mutants of H3.3 and studied its effects on chromatin.

114

Genes with human-specific features, ncRNAs and innovations in genetic and epigenetic mechanisms cooperate to brain evolution

Maina Bitar[1], Elizabeth O'Brien[1], Stefanie Kuiper[1], Guy Barry[1]

1. QIMR Berghofer, Brisbane / Herston, QUEENSLAND, Australia

Humans have adapted to widespread changes during the past two million years in both environmental and lifestyle factors but our knowledge of how the human brain differs from those of other species in terms of evolutionary adaptations and unique functionality is still very limited. Comparative genomics studies reveal valuable insight, and help determine variations in genomic sequence that may provide functional information to better understand species-specific adaptations. Recent discoveries (e.g. RNA modifications and editing, circular RNAs and multiple long non-coding RNA encoded functions) demonstrate our emerging knowledge of evolving brain function. One of the main routes to generate human-specific features is trough genomic changes, which are therefore valuable in uncovering genes and pathways contributing to development and function, especially in the brain. We updated the list of high confidence human-specific genomic variants that associate with protein-coding genes and found 845 such regions. These are mostly due to gene duplications, the emergence of novel human genes and sequence and structural alterations. Network analysis of these associated protein-coding genes identifies adaptations to brain, immune and metabolic functions as highly involved. We further show that many of these protein-coding genes may be functionally associated with neural activity and generating the expanded human cortex in dynamic spatial and temporal contexts. Our results are consistent with environmental changes, such as immune challenges and alterations in diet, as well as neural sophistication, as significant contributors to recent human evolution. These new findings present an opportunity to combine newly discovered mechanisms, both genetic and epigenetic, with more established concepts into an improved and inclusive picture to better understand the uniquely evolved human brain.

115

msgbsR: An R package for analysing methylation-sensitive restriction enzyme sequencing data

Benjamin T Mayne[1, 2], Sam Buckberry[3, 4], Shalem Y Leemaqz[1], Carlos Rodriguez Lopez[5], Claire T Roberts[1], Tina Bianco-Miotto[1, 5], Jimmy Breen[1, 6]

1. Robinson Reseach Institute, University of Adelaide, Adelaide, SA, Australia

2. Cancer Research, South Australian Health & Medical Research Institute (SAHMRI), Adelaide, SA, Australia

3. Harry Perkins Institute of Medical Research, University of Western Australia, Perth, WA, Australia

4. Plant Energy Biology, ARC Centre of Excellence, University of Western Australia, Perth, WA, Australia

5. School of Agriculture, Food and Wine, University of Adelaide, Adelaide, SA, Australia

6. Bioinformatics Hub, University of Adelaide, Adelaide, SA, Australia

Genotyping-by-sequencing (GBS) or restriction-site associated DNA marker sequencing (RAD-seq) is a practical and cost-effective method for analysing large genomes from high diversity species. This method of sequencing, coupled with methylation-sensitive enzymes (often referred to as methylation-sensitive restriction enzyme sequencing or MRE-seq), is an effective tool to study DNA methylation in parts of the genome that are inaccessible in other sequencing techniques or are not annotated in microarray technologies. Current software tools do not fulfil all methylation-sensitive restriction sequencing assays for determining differences in DNA methylation between samples. To fill this computational need, we present *msgbsR*, an R package that contains tools for the analysis of methylation-sensitive restriction enzyme sequencing experiments. *msgbsR* can be used to identify and quantify read counts at methylated sites directly from alignment files (BAM files) and enables verification of restriction enzyme cut sites with the correct recognition sequence of the individual enzyme. In addition, *msgbsR* assesses DNA methylation based on read coverage, similar to RNA sequencing experiments, rather than methylation proportion and is a useful tool in analysing differential methylation on large populations. The package is fully documented and available freely online as a Bioconductor package (https://bioconductor.org/packages/release/bioc/html/msgbsR.html).

116

VIRAL GENOME ANALYSIS: BIOINFORMATIC ANALYSIS OF RNA ELEMENTS ESSENTIAL FOR REPLICATION.

Christopher M Brown[1], David Chyou[1], Stephan Roessner[1], Chun Shen Lim[1]

1. University of Otago, Dunedin, Otago, New Zealand

Many pathogenic viruses have RNA genomes (e.g. influenza, Hepatitis C virus) and as expected these contain all the required elements for the viral lifecycles. These elements may be structured or unstructured. In addition, viruses with DNA genomes often have key regulatory RNA elements in their genomes (e.g. HIV, hepatitis B virus).

Here we report approaches we are developing and using to discover novel functional elements in viral genomes. These have been applied to model the functional elements in HBV. Notably, HBV contains a complex post-transcriptional regulatory element that inhibits splicing and promotes the export of unspliced RNA. This element has rtecently been shown to be able to be targeted by antiviral drugs. This element contains a combination of structured and unstructured elements that regulate this process.

We are also investigating the roles of RNA-RNA interactions in the CRISPR-Cas antiviral response. To facilitate this we have developed publicly available software available through bioanalysis.otago.ac.nz

1. Lim, Chun Shen.& Brown, C.M. (2018) Know Your Enemy: Successful Bioinformatic Approaches to Predict Functional RNA Structures in Viral RNAs. Frontiers in Microbiology, 8:2582
2. Lim, Chun Shen & Brown, C.M. (2016) Hepatitis B virus nuclear export elements: RNA stem-loop alpha and beta, key parts of the HBV post-transcriptional regulatory element. RNA Biol, 13:743

117

Links between DNA replication stress and telomere extension in cancer cells

Tracy M Bryan[1], Melissa Kartawinata[1], Adrian S Tong[1], Josh L Stern[1], Amanda XY Chen[1]

1. Children's Medical Research Institute, Westmead, NSW, Australia

The ribonucleoprotein enzyme telomerase counteracts telomere shortening in germ cells, stem cells and a majority of cancers, allowing for unlimited cell division. Most normal cells do not express telomerase, so

inhibition of telomerase is a promising avenue for development of specific anti-cancer treatments that should be applicable to almost all cancers. Telomerase action can be inhibited *in vivo* by blocking the cellular pathways that direct telomerase to chromosome ends, i.e. telomerase 'recruitment' to telomeres. This process is highly regulated, but factors that regulate it are incompletely understood.

We have demonstrated that telomerase recruitment to telomeres in human cells is highly dependent on the DNA damage response kinases ATM and ATR[1], that together play a crucial role in safeguarding the genome. This is an unexpected finding, since in general the DNA damage response is suppressed at telomeres, suggesting that telomeres exploit DNA damage signalling pathways to bring telomerase to the telomere. ATR controls the cellular response to single-stranded DNA arising during replication fork stalling, and we also showed that replication stress increases the levels of telomerase at the telomere, in an ATR-dependent manner[1]. The effect of ATR on telomerase recruitment is mediated, at least in part, by its substrate Chk1, a kinase that is also critical for genomic stability during DNA replication.

We have also elucidated the downstream targets of ATM that are crucial for telomerase recruitment, including the telomeric "shelterin" proteins TPP1 and TRF1, which are responsible for positively or negatively regulating interactions of telomerase with telomeres, respectively. Together, these observations represent a new understanding of the relationship between telomerase and the DNA damage response, revealing telomerase as a specialised response to general genomic stress. Furthermore, both processes are key drivers of oncogenesis, and our data demonstrate that they are tightly linked.

1. Tong, A.S., Stern, J.L, Sfeir, A., Kartawinata, M., de Lange, T., Zhu, X.D. and Bryan, T.M. (2015) ATM and ATR Signaling Regulate the Recruitment of Human Telomerase to Telomeres. Cell Reports, 13(8):1633-46.

118

Analysis of functionally deficient variants of transcription factor *SIM2* present in patients with intellectual disabilities

Emily L Button[1], David C Bersten[1], Murray L Whitelaw[1]

1. University of Adelaide, Adelaide, SA, Australia

Single-Minded 2 (SIM2) is a member of the basic-helix-loop-helix/PER-ARNT-SIM (bHLH/PAS) family of transcription factors. These proteins are known to play important roles in biological processes such as development, oxygen homeostasis and stress responses, however the physiological role of SIM2 is largely unknown. SIM2 functions as a heterodimer with ARNT2 to regulate gene expression, and can act as both a transcriptional repressor and activator in a highly context dependant manner. *Sim2* mRNA is expressed within the brain both during development and postnatally, however the function neuronal function of SIM2 is not well characterised. *Sim2* knockout (KO) mice die perinatally, displaying varying phenotypes including dysmorphologies such as cleft palates and sporadic scoliosis. These mice also display a reduced number of somatostatin and thyrotropin releasing hormone expressing neurons within the hypothalamus. A screen was performed on an exome sequencing database to identify *SIM2* variants in patients with intellectual disabilities, delayed development of speech, dysmorphic features and scoliosis. We performed reporter gene assays on a number of the variants and identified non-synonymous mutations that cause a change in transcriptional activity. These variants were then further characterised in order to determine the molecular mechanism behind this change in activity. This study highlighted several *SIM2* variants found in patients with disabilities and validated a candidate set as potential disease causing or contributing variants.

119

Antisense oligonucleotide-mediated exon skipping: a potential therapeutic strategy for Marfan syndrome

Jessica Cale[1, 3, 2], Kane Greer[3, 2], Sue Fletcher[3, 2], Steve Wilton[3, 2]

1. Murdoch University, Murdoch, WA, Australia

2. Centre for Comparative Genomics, Murdoch University, Murdoch, WA, Australia

3. Perron Institute for Neurological and Translational Sciences, Perth, WA, Australia

Marfan syndrome (MFS) is a dominant connective-tissue disorder caused by *fibrillin-1* (*FBN1*) mutations. Fibrillin-1, is a large glycoprotein that aggregates into multimers to form the backbone of microfibrils, providing structural and regulatory support to connective tissues. MFS shows variable age of onset and is progressive in nature, with major clinical manifestations in the skeletal, ocular and cardiovascular systems. The aetiology of MFS is not fully understood, however, mutations in *FBN1* are thought to result in heterogeneous fibrillin-1 monomers that are unable to form multimers, leading to reduction in microfibril abundance. To date, over 1,800

unique mutations are reported in *FBN1*, including a c.6354C>T mutation resulting in exon 52 skipping from ~half the processed mRNA transcripts. This study demonstrates a potential therapeutic strategy for MFS using antisense oligonucleotides to alter the exon structure, of both the normal and disease causing mRNA transcripts, to re-establish the periodicity of fibrillin-1. We hypothesise that the resulting modified fibrillin-1 monomers will correctly form multimers, increasing microfibril abundance and reducing disease severity.

Antisense oligonucleotides, designed to target regulatory splicing motifs within *FBN1* exon 52, were screened in normal and patient fibroblasts, to assess induced exon 52 skipping from the mRNA transcripts. Treated cells were also immunostained using an anti-fibrillin-1 antibody to reveal the abundance and morphology of microfibrils. The observed exon 52 skipping was dose dependant, with up to 99% of transcripts in patient fibroblasts missing exon 52. A corresponding increase in microfibril abundance was observed in treated, compared to untreated patient cells

The use of antisense oligonucleotides to induce targeted alternative splicing has garnered attention in recent years, particularly for the treatment of Duchenne muscular dystrophy. We believe this technique is applicable to MFS, with preliminary *in vitro* data supporting the hypothesis that inducing homogeneity between fibrillin-1 monomers is a potential therapeutic option.

120

Efficient homology-directed repair using single-stranded DNA templates

Michel Cannieux[1]

1. Integrated DNA Technologies (IDT), Baulkham Hills, NSW, Australia

Homology-directed repair (HDR) is a valuable tool for genome editing when combined with the CRISPR/Cas9 system. HDR with a targeted repair template after a double-stranded DNA (dsDNA) break enables the generation of knock-in strains or animal models, including insertion of reporter genes or the introduction of specific mutations. Double-stranded DNA can be used as a repair template, however knock-in efficiency can be low. Further, linear double-stranded templates can insert at any break present in the genome, including off-target sites. Several recent studies have demonstrated that single-stranded DNAs (ssDNAs) are superior HDR templates. ssDNAs can be obtained as synthetic oligonucleotides or made from longer dsDNA templates. Here, we report optimal conditions to make point mutations and epitope tag insertions using IDT Ultramer® Oligonucleotide ssDNA HDR templates delivered with a Cas9 ribonucleoprotein complex. Investigations of strand choice (targeting vs. non-targeting strand), homology arm length, strand symmetry, Cas9 variants (wildtype, nickase, HiFi), chemical modification, and Ultramer purification conditions are shown and a rule set for HDR using oligonucleotide-based templates of 200 bases or less is presented. IDT Megamer™ ssDNA fragments are used for insertions of 200-2000 bases. Initial experiments suggest that these long ssDNA HDR templates show similar advantages and behavior as their shorter oligonucleotide counterparts. Furthermore, Megamer ssDNA fragments injected with Cas9 ribonucleoprotein complexes into mouse zygotes successfully produce correctly targeted alleles in live offspring.

121

Disrupting promoter-enhancer interaction of *Bcl11b* in T cell by CRISPR/Cas9

Wing Fuk Chan[1], **Timothy Johanson**[1], **Hannah Coughlan**[1], **Gordon Smyth**[1], **Rhys Allan**[1]

1. Walter and Eliza Hall Institute, Parkville, VIC, Australia

Recent advances in understanding the 3D structure of chromosomes has led to increased focus on the role of genome architecture in regulating gene transcription. Above the nucleosomal level, DNA elements like a gene promoter can form loop with distal enhancer element to drive gene transcription. Such loop can be lineage specific and is thus believed to play a crucial role during immune cell development and differentiation. However, how such loop is established and maintained is poorly understood.

Bcl11b is a critical transcription factor during T cell development, it is needed for the proper development of T lymphocytes and the deletion of it can skew the developing thymocytes towards natural killer or myeloid lineages. Here the genomic architecture of *Bcl11b* in T and B lymphocytes was investigated using chromosome conformation capture coupled with high-throughput sequencing (Hi-C) and CRISPR/Cas9 genome editing technologies.

Hi-C data of T and B lymphocytes on the *Bcl11b* regions revealed a T cell specific interaction between the *Bcl11b* promoter and a distal (~850 kb) cis-element 3' downstream, the location of which is consistent with the previously characterised enhancer element of *Bcl11b*. By exploiting the CRISPR/Cas9 gene editing technology, the enhancer was bi-allelically deleted in the T cell lymphoma cell line EL4 and the expression of *Bcl11b* was found to be down-regulated but not abolished.

This preliminary work provides a framework of how a cis-regulatory region of the genome can be manipulated and this approach would further allow us to understand the role of the cistrome during immune development and differentiation.

122

Differential methylation analysis of reduced representation bisulfite sequencing experiments

Yunshun Chen[1, 2], Bhupinder Pal[1, 2], Jane Visvader[1, 2], Gordon Smyth[1, 3]

1. Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia

2. Department of Medical Biology, The University of Melbourne, Parkville, VIC, Australia

3. Department of Mathematics and Statistics, The University of Melbourne, Parkville, VIC, Australia

Studies in epigenetics have shown that DNA methylation is a key factor in regulating gene expression. DNA methylation typically occurs in CpG context. When located in a gene promoter, DNA methylation often acts to repress transcription and gene expression. The most commonly used technology of studying DNA methylation is bisulfite sequencing (BS-seq), which can be used to measure genomewide methylation levels on the single-nucleotide scale. Notably, BS-seq can also be combined with enrichment strategies such as reduced representation bisulfite sequencing (RRBS) to target CpG-rich regions in order to save per-sample costs.

A typical DNA methylation analysis often involves identifying differentially methylated regions (DMRs) between different experimental conditions. Many statistical methods have been developed for finding DMRs in BS-seq data. In this talk, I will describe a novel approach of detecting DMRs using edgeR. A case study will be provided to demonstrate how differential methylation analyses can be fit into the existing pipelines specifically designed for RNA-seq differential expression studies. The method proposed in the talk can be applied to any BS-seq data that includes some replication, but it is especially appropriate for RRBS data with small numbers of biological replicates.

123

Identification of ALT phenotypes in telomerase positive cancers

Eugene Choi[1], Joyce Lee[1], Hilda Pickett[1], Roger Reddel[1]

1. Children's Medical Research Institute, Westmead, NEW SOUTH WALES, Australia

The immortality characteristics of the cancer cells is contributed by the maintenance of telomere lengths. Telomerase, a protein involved in elongation of the telomeres occurs in approximately 80% of cancers (1). The remaining cancers involve a mechanism which is independent of telomerase known as Alternative Lengthening of Telomeres (ALT) and Ever Shortening Telomeres (EST). A preliminary screening for ALT (C-circles) and Telomerase (qTRAP) in 384 cancer cell lines was performed. We discovered a few cancer cell lines with dual-telomere maintenance mechanism (Dual-TMM) indicated by the low levels of C-circles and telomerase activity. Further identification of the ALT characteristics in the cell lines were showed promising results that this is the case.

1. Bryan TM, Englezou A, Dalla-Pozza L, Dunham MA, Reddel RR. Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. Nat Med. 1997;3:1271-4.

124

Non-coding transcription and mitotic recombination of the ribosomal RNA gene repeats in cancer

Daria A. Chudakova[1, 2], Jessica L. Fitch[3, 1], Megan A. Schischka[1], Katherine M. Hannan[4], Ross D. Hannan[4], Austen R.D. Ganley[1, 2]

1. School of Biological Sciences, University of Auckland, Auckland, New Zealand

2. Maurice Wilkins Centre for Molecular Biodiscovery, University of Auckland, Auckland, New Zealand

3. Institute of Natural and Mathematical Sciences, Massey University, Albany, Auckland, New Zealand

4. Department of Cancer Biology and Therapeutics, Australian National University, Canberra, Australia

Ribosomal RNAs – the structural and catalytic components of ribosomes – are encoded by multiple copies of head-to-tail tandemly arrayed ribosomal RNA genes (rDNA) separated by an intergenic spacer (*IGS*). The rDNA is a recombination hotspot, and as a consequence of this and its repetitive nature, is one of the most unstable regions of the genome. Interestingly, several types of solid tumours have chromosomal rearrangements involving the rDNA. In addition, a number of non-coding RNAs (ncRNAs) are transcribed from the IGS region of rDNA, and our preliminary data indicates that their transcription is increased during malignant

progression in the Eu-Myc mouse lymphoma model. However, the role of rDNA recombination and non-coding transcription in tumorigenesis is yet to be investigated.

Here we performed hybridisation capture RNA-sequencing to determine the full profile of long ncRNAs transcribed from the rDNA IGS in Eu-Myc cells. From this, we identified a number of ncRNAs located across the IGS. We are also developing a chromatin immunoprecipitation-based assay (ChIP-qPCR) to measure rDNA recombination in mammalian cells, based on assumption that recombination factors (such as Rad50, Rad54 and RBBP8) are recruited to DNA in proportion to the level of recombination. We have shown that recombination factors are detectable at rDNA by ChIP-qPCR and, in conditions where the recombination rate is expected to differ, their levels vary consistently. Identification of the ncRNAs transcribed from the rDNA and development of a tool to measure rDNA recombination will provide a platform for determining whether altered recombination and non-coding transcription in the rDNA play roles in malignancy.

125

Fine-mapping reveals complex genetic architecture underlying DNA methylation

Vignesh Kartik Chundru[1], Riccardo E Marioni[2], Ian J Deary[2], Naomi R Wray[1], Peter M Visscher[1], Allan F McRae[1]

1. Institute for Molecular Biosciences, University of Queensland, Brisbane, QLD, Australia

2. Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, United Kingdom

The genetic contribution to DNA methylation has been studied through identifying methylation quantitative trait loci (mQTL). A trivial example of an mQTL is a SNP at the CpG site (CpG-SNP) which will change the C or G in the CpG sequence, and consequently the DNA methylation. Such sites have a known causal variant with a large effect and thus are potential model traits to test fine-mapping methodology. Using DNA methylation data from the Lothian Birth Cohorts of 1921 and 1936 (n=1366), we compare two methods never previously used for fine-mapping (the J-test and BSLMM) with the widely used BIMBAM method by constructing a 95% credible set of causal SNPs for mQTL with a CpG-SNP. All three methods failed to capture the CpG-SNP with the expected 95% probability. Simulations confirmed the three methods performed as expected under ideal conditions. The influence of multiple, independent signals underlying the mQTL was also excluded. Comparing the CpG-SNP, and the most associated SNP (top-SNP), we found that often the effect of the top-SNP was masking the effect of the CpG-SNP. This was shown to be more evident in CpG islands, where the fine-mapping methods captured the CpG-SNP less often than in non-island regions (OR=1.6, $p=2\times10^{-3}$). This indicates that methylation sites within a CpG island and the surrounding region can share common genetic control, providing potential insight into the mechanisms underlying the maintenance of DNA methylation at CpG islands.

126

Long-read sequencing reveals the splicing profile of the neuropsychiatric disease gene *CACNA1C* in human brain

Michael B Clark, Tomasz Wrzesinski, Aintzane B Garcia, Paul J Harrison, Joel E Kleinman, Thomas Hyde, Daniel R Weinberger, Wilfried Haerty, Elizabeth M Tunbridge

RNA splicing is a key mechanism linking genetic variation and complex diseases, including schizophrenia. Splicing profiles are particularly diverse in the brain, but it can be difficult to accurately identify and quantify full-length mRNA isoforms using standard approaches, impeding the identification of splice isoforms linked to disease. Emblematic of this is *CACNA1C*, a large gene that shows robust genetic associations with several psychiatric disorders and encodes multiple, functionally-distinct, voltage-gated calcium channels via alternative splicing. However, the extreme length (>13kb), number of exons (>50) and high level of alternative splicing means *CACNA1C* isoform structure and expression is poorly understood.

We have investigated *CACNA1C* expression in six regions of post-mortem human brain by long-read Nanopore sequencing of the complete coding sequence from expressed transcripts. The ability of Nanopore sequencing to define the complete exonic structure of *CACNA1C* transcripts allows us to identify known and novel gene isoforms, including novel exons, microexons and multi-exonic skipping events and their expression across multiple brain regions. We show that the *CACNA1C* splice isoform profile varies between brain regions and is substantially more complex than currently appreciated: we identified 34 novel exons and 83 high confidence novel transcripts, a number of which are abundantly expressed and predicted to alter protein function. Our findings demonstrate that knowledge of human splice isoform diversity remains far from complete. Furthermore, the accurate characterisation of *CACNA1C* isoforms with long-read sequencing will now facilitate the identification of disease-linked isoforms and future studies on *CACNA1C*'s role in psychiatric disorders.

127

Electron microscopy of human telomerase

Scott Cohen[1], Rosalba Rothnagel[2], Tram Phan[3], George Lovrecz[3], Tim Adams[3], Tracy Bryan[1], Michael Parker[4], Ben Hankamer[2]

1. Children's Medical Research Institute, Westmead, NSW, Australia

2. Institute for Molecular Bioscience, University of Queensland, St. Lucia, QLD, Australia

3. CSIRO Manufacturing, Parkville, VIC, Australia

4. St. Vincent's Institute for Medical Research, Fitzroy, VIC, Australia

Telomeres, the repetitive DNA-protein complexes at the ends of linear chromosomes, shorten with each cycle of DNA replication, providing a counting mechanism to limit the number of times a cell can divide. Most cancer cells have activated the ribonucleoprotein enzyme **telomerase** to add telomeric DNA repeats and counteract telomere shortening, allowing for unlimited proliferation. Although inhibition of telomerase has been considered a promising approach to cancer therapy for more than two decades, its low cellular abundance (~50-100 copies/cell) and challenging biochemistry have stymied development of small-molecule inhibitors.

We reported the purification and composition of the core human telomerase enzyme complex, consisting of two molecules each of: i) the telomerase reverse transcriptase catalytic protein; ii) telomerase RNA; and iii) the RNA-binding protein dyskerin (*1*). Building on this knowledge we developed an over-expression system in suspension HEK-293T cells that yields ~400-fold greater activity over endogenous levels; this system is providing sufficient telomerase for electron microscopy studies. We have determined a low-resolution structure by negative-stain EM, revealing an elongated, bilobal structure. We are currently in the early stages of imaging with cryo-EM, experimenting with functionalised derivatives of graphene to immobilise particles on the grid before vitrification. Our long-term aim is to apply structure-guided design to the development of small-molecule telomerase inhibitors.

(*1*) Cohen SB, et al. (2007) *Science*, 315, pp 1850-1853.

128

Chiron: Translating nanopore raw signal directly into nucleotide sequence using deep learning

Haotian Teng[1], Minh Cao[1], Michael Hall[1], Tania Duarte[1], Lachlan Coin[1]

1. Institute for Molecular Bioscience, St Lucia, QLD, Australia

Sequencing by translocating DNA fragments through an array of nanopores is a rapidly maturing technology which offers faster and cheaper sequencing than other approaches. However, accurately deciphering the DNA sequence from the noisy and complex electrical signal is challenging. Here, we report Chiron, the first deep learning model to achieve end-to-end basecalling: directly translating the raw signal to DNA sequence without the error-prone segmentation step. Trained with only a small set of 4000 reads, we show that our model provides state-of-the-art basecalling accuracy even on previously unseen species. Chiron achieves basecalling speeds of over 2000 bases per second using desktop computer graphics processing units, making it competitive with other deep-learning basecalling algorithms.

129

Structural and functional analysis of the *C. elegans* TRIM protein NHL-2

Rhys N Colson[1], Peter R Boag[1], Jacqueline A Wilce[1]

1. Biochemistry and Molecular Biology, Monash Biomedicine Discovery Institute , Melbourne, VIC, Australia

TRIM-NHL proteins play a role in various cellular functions from muscular and neuronal tissue development to controlling stem cell fate. While the NHL domain of these proteins was originally thought to facilitate protein-protein interactions, it has recently been discovered that several TRIM-NHL proteins have sequence specific interactions with RNA through the domain. The *C. elegans* protein NHL-2 is involved in small RNA pathways, interacting with the DEAD-box RNA helicase's CGH-1 and DRH-3. I have shown that, like other TRIM-NHL proteins, NHL-2 binds directly to U-rich RNA and identified residues key to RNA binding, which when mutated show a significant reduction in affinity. These mutations will be tested in vivo, where the phenotypes of RNA binding mutant worms will be compared to *nhl-2* knock out, assessing the biological importance of NHL-2's RNA binding.

Structural analysis of NHL-2 is also being undertaken, with aim to use using X-ray crystallography to solve the structure of the NHL domain both apo and in complex with RNA. Currently the only solved structure of the NHL

domain is from the *drosophila* protein BRAT. Solving the structure of NHL-2's NHL domain will provide insight into the structural differences between TRIM-NHL proteins that underlie their functional differences. SAXS is also being used to investigate the quaternary structure of NHL-2, looking at the way in which homodimerization, that is facilitated by the coiled coil domain, affects the positioning of other domains.

130

Mutation Analysis and Effect on Prognosis of Direct Antiglobulin Test Positivity in Diffuse Large B-Cell Lymphoma

Caitlin Coombes[1, 2], Joshua Tobin[3, 4], Kartik Saxena[2], Jun Hee Lim[5], Bruce Shadbolt[6], Yogesh Jeelall[7, 8], Chris Goodnow[7, 9], Thomas D Andrews[7], Sanjiv Jain[10], Simon Jiang[7, 11], Keisuke Horikawa[5], Dipti Talaulikar[2, 5]

1. Australian National University, Acton, ACT, Australia

2. Haematology Translational Research Unit, Haematology Department, The Canberra Hospital, Garran, ACT, Australia

3. Princess Alexandra Hospital, Brisbane, QLD, Australia

4. Diamantina Institute, University of Queensland, Brisbane, QLD, Australia

5. Department of Cancer Biology and Therapeutics, The John Curtin School of Medical Research, Australian National University, Acton, ACT, Australia

6. Centre for Advances in Epidemiology and IT, The Canberra Hospital, Garran, ACT, Australia

7. Department of Immunology and Infectious Disease, The John Curtin School of Medical Research, Australian National University, Acton, ACT, Australia

8. School of Medicine and Pharmacology, Harry Perkins Institute of Medical Research, The University of Western Australia, Perth, WA, Australia

9. Department of Immunology , Garvan Institute of Medical Research, Darlinghurst, NSW, Australia

10. Anatomical Pathology Department, The Canberra Hospital, Garran, ACT, Australia

11. Renal Medicine Department, The Canberra Hospital, Garran, ACT, Australia

Introduction

Direct Antiglobulin Test (DAT) positivity occurs at higher rates in patients with Non-Hodgkin Lymphoma. This study examined the genetic differences between DAT positive (DAT +) and DAT negative (DAT-) DLBCL cases, and the effect of DAT positivity on survival and revised international prognostic index (R-IPI) in Diffuse Large B-Cell Lymphoma (DLBCL).

Method

One hundred and thirteen cases that had undergone DAT testing at the time of DLBCL diagnosis were included. DNA was extracted from 7 DAT+ and 8 DAT- patient samples and targeted next-generation sequencing (NGS) was performed. Analysis of survival outcomes was assessed by the Kaplan-Meier method and Cox proportional hazards regression.

Results

An average of 10.07 mutations per DLBCL clone were found on NGS (DAT+ 10.14 vs DAT-10.00). Mutations were found in genes and B-cell pathways associated with oncogenesis and autoimmunity in both DAT+ and DAT- groups. The most recurrently mutated genes were *KMT2D* (n=13), *MYOM2* (n=9), and *EP300* (n=8), which were all mutated in both groups. The genes associated with the non-canonical NFκB pathway were found to be more frequently mutated in the DAT+ group, with 6 mutations in the DAT+ group compared with 1 mutation found in the DAT- group.

There was a non-significant difference in overall survival between the DAT+ and DAT- groups (p=0.19). DAT significantly influenced the effect of R-IPI on survival (p=0.017). Within the DAT- group, R-IPI maintained its strong predictive power for overall survival (p<0.0001). Within the DAT+ group, the two represented R-IPI groups were not significantly different in their overall survival (p=0.74).

Conclusion

Both DAT positive and DAT negative patient groups showed high rates of non-synonymous mutations in genes that are frequently targeted in DLBCL. This study did not demonstrate a significant effect on survival of DAT positivity in patients with DLBCL, but a moderating influence on R-IPI's survival effect.

131

Four gene regulatory functions of oncogenic potential for NONO protein in neuroblastoma

Jack A L Cooper[1], Alina Naveed[1], Ryan Lister[2], Archa Fox[1, 2]

1. School of Human Sciences, The University of Western Australia, Perth, WA, Australia

2. School of Molecular Sciences, The University of Western Australia, Perth, WA, Australia

Neuroblastoma is a childhood cancer characterised by complex, heterogeneous genetics and gene regulation. The long-term survival rate of high risk neuroblastoma patients is less than 50%, due in large part to the lack of targeted treatment. Neuroblastomas show few recurrent genetic mutations. Instead, it has been suggested that the cancer is driven by aberrant activity of the numerous gene regulators which are active in the development of neural-derived tissue. One such gene regulator is NONO, a protein whose expression has been strongly associated with poor prognosis in neuroblastoma. However, no comprehensive mechanism has yet been offered to explain this connection. By integrating PAR-CLIP, ChIP and NONO knockdown RNA sequencing datasets, we have obtained evidence for four gene regulatory functions for NONO with connections to tumorigenesis. Firstly, NONO binds to pre-mRNA transcripts to decrease their abundance. Secondly, NONO directs alternate splicing, and appears to promote the formation of a truncated transcript of the ALK oncogene. Thirdly, NONO modulates the activity of the CREB1 transcription factor, whose dysregulation has previously been associated with multiple cancers. Fourthly, NONO indirectly upregulates the targets of cholesterogenic transcription factor SREBP1A, enhancing cholesterol synthesis – a biological process whose upregulation is prognostic for poor outcome in neuroblastoma. As the first study to comprehensively characterise NONO's activity in neuroblastoma, we lay the foundations for targeted investigation of NONO's potential as a therapeutic target. Future work will confirm the importance of each of NONO's functions to its association with neuroblastoma and other cancers.

132

Detecting RNA Structure Heterogeneity in Cells

Vincent Corbin[1, 2, 3], Phillip Tomezsko[4, 3], Paromita Gupta[3], Margalit R Glasgow[3], Sitara Persad[3], Lachlan McIntosh[1, 2], Tony Papenfuss[1, 2], Silvia Rouskin[3]

1. Bioinformatics, Walter and Eliza Hall Institute, Parkville, VIC, Australia

2. University of Melbourne, Melbourne, VIC, Australia

3. Whitehead Institute of Biomedical Research, Cambridge, MA, USA

4. Harvard University, Cambridge, MA, USA

RNA is unique among all biomolecules as it can be both information storing and enzymatic. Indeed, in addition to the genetic code, RNA possesses a second layer of information integrated in its secondary structure that can regulate processes as varied as splicing, localization, translation efficiency and protein binding. However, our current ability to explore the complexity of RNA structure is limited. Prediction algorithms do not account for intra-cellular interactions such as the role played by proteins in RNA folding or dynamic unwinding of structured regions by ribosomes. Other existing assays based on RNA chemical probing (DMS, SHAPE) can explore RNA structure inside the cell but fail to untangle the multiple conformations RNA may assume inside a single cell, leading to a potentially false interpretation of its structure. We developped a method that is capable of detecting alternative RNA structures which form from the same underlying sequence, both in vitro and ex vivo. We applied our approach to HIV-1 and revealed genomic RNA structure heterogeneity with novel functional implications.

133

FGF9 homodimerization is required for male sex determination

Brittany Croft, Anthony Bird, Makoto Ono, Liang Zhao, Masayo Harada, Andrew Sinclair, Vincent Harley

Disorders of sex development (DSDs) are a range of congenital conditions, including 46,XY gonadal dysgenesis (GD), where only 50% of cases receive a specific molecular diagnosis. Improving our understanding of the genetic causes of 46,XY GD is critical to improve clinical diagnosis and management of these conditions. Among the genes promoting male sex determination, is *fibroblast growth factor (FGF) 9*.

FGF9 is critical for the repression of pro-ovarian signalling pathways such as WNT4/RSPO1 and FOXL2. This ensures sufficient levels of SOX9 are expressed in the somatic cells of the embryonic gonad to drive Sertoli cell differentiation and ultimately, male testicular development. Loss of Fgf9 results in complete male to female

sex reversal, while Fgf9 gene mutants causing impaired FGF9 homodimerization in humans and mice lead to skeletal defects such as synostosis.

Here we investigate the requirement of FGF9 dimer formation for testicular development. Using homodimer-compromised FGF9 mutants from both a mouse model and a human DSD patient.

The spontaneous mouse Fgf9 mutant; Elbow knee synostosis (Eks), amino acid substitution (N143T), which lies at the homodimerization interface results in impaired FGF9 homodimerization. Examination of XY Fgf9N143T/N143T gonads showed delayed testes cord development and ectopic expression of the female Granulosa cell marker FOXL2 at the gonadal poles, indicative of XY sex reversal.

We have identified a 46,XY GD DSD patient with an amino acid substitution (D195N), previous studies indicated that the D195 residue is critical for the homodimerization of FGF9. Purified recombinant FGF9D195N protein showed reduced affinity for heparin, a property required for homodimerization. *In vitro* analysis showed reduced ability to induce Sertoli cell proliferation, which is required for normal testis development.

In summary, this data suggests that homodimerization of FGF9 is critical for FGF-mediated male sex determination.

134

Can genomic and phenotypic analyses unravel how the yeast Saccharomyces cerevisiae evolves in emulsion culture?

Emelyne Cunnington[1], Thomas Pfeiffer[1]

1. Massey University, Albany, AUCKLAND, New Zealand

A trade-off, in biology, is defined as a constraint that link two traits. Improving one trait has a negative impact on the other trait. While there is an increased interest in trade-offs, and their roles in shaping the traits of organisms, it remains difficult to demonstrate their presence and relevance in lab experiments. To study a trade-off between yield (biomass production per resource) and rate (biomass production over time), we conducted an experimental evolution study with Saccharomyces cerevisiae in an emulsion culture that has previously been shown to select for yield (Bachmann *et al*, 2013). We observed that 5 of our 6 lines increased in growth yield during evolution. While this increase was not correlated with a decrease in growth rate, our study allowed us to uncover other adaptations. Genomic analysis of the ancestral and evolved lines provided further insights into the evolutionary response of *S. cerevisiae* to selection in the emulsion environment.

1. Bachmann, H. et al. Availability of public goods shapes the evolution of competing metabolic strategies. Proc. Natl. Acad. Sci. U. S. A. 110, 14302–7 (2013).

135

Mechanism for RNA-mediated regulation of the histone methyltransferase Polycomb repressive complex 2 (PRC2)

Chen Davidovich[2, 1]

1. Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria 3800, Australia

2. EMBL-Australia and the ARC Centre of Excellence in Advanced Molecular Imaging, Clayton, Victoria 3800, Australia

Interactions between epigenetic modifiers and long noncoding RNAs or mRNAs were widely reported and were linked to normal development and to disease. Yet, functional studies to elucidate how, or if, such epigenetic modifiers are regulated by their RNA binding partners are often complicated by promiscuous RNA binding. From the protein side, lack of canonical RNA recognition motifs (RRMs) or sequence homology to other RNA binding proteins often complicates mechanistic studies.

The Polycomb repressive complex 2 (PRC2) is a histone methyltransferase that trimethylates K27 of histone H3 (H3K27me3). PRC2 is essential for the maintenance of the repressed epigenetic state of thousands of genes during development and is dysregulated during cancer. PRC2 binds thousands of transcripts, but despite a decade of extensive study a PRC2-binding motif within its target RNAs was not identified. Furthermore, an RNA-binding domain within PRC2 was not detected and canonical RRMs are not present within any of its protein subunits. Collectively, these complicated mechanistic studies to determine the function of PRC2-RNA interactions.

Through combining biochemical and biophysical approaches with genomics and mass-spectrometry we identified the PRC2-binding motif within RNA and the RNA-binding domain within PRC2. Multiple short tracts of consecutive guanines largely increase the affinity of PRC2 to RNA and significantly associate with PRC2 binding sites on RNA transcripts in cells. DNA sequences coding for PRC2-binding RNA motifs are enriched at PRC2-binding sites on chromatin and H3K27me3-modified nucleosomes, provide means for RNA-mediated

regulation of PRC2 *in cis*. Additional new data will be discussed from our on-going study to understand how the polycomb machinery is regulated.

136

Using *in utero* electroporation to model a second-hit mechanism for GATOR1-related focal epilepsy in mice

Ruby Dawson[1], James Hughes[1], Melissa White[1, 2], Sandra Piltz[1, 2], Paul Thomas[1, 3]

1. University of Adelaide, Adelaide, SA, Australia

2. SA Genome Editing , Adelaide , SA, Australia

3. South Australian Health and Medical Research Institute (SAHMRI), Adelaide, SA, Australia

Mutations in *DEPDC5*, *NPRL2* and *NPRL3*, which make up the GATOR1 complex, have been found in a cohort of families with focal epilepsy. However the mechanisms of how these cause the disease remains elusive. In vitro studies have shown that GATOR1 functions to inhibit mTOR signalling and hyperactivity of this pathway has independently been linked with epilepsy. mTOR dysregulation is therefore hypothesised to be the major factor in the pathology of GATOR1-related epilepsy. We have developed a functional assay using CRISPR null cell lines where the null phenotype of hyperactive mTOR can be rescued by expressing the wildtype protein. GATOR1 mutations found in patients can therefore be screened for loss-of-function in the context of mTOR. Multiple germline mutations have been confirmed to have lost this function partially or completely. Somatic GATOR1 mutations have also been identified in patients and can be screened to investigate a 'second-hit' mechanism of disease, where seizures are proposed to result from a second somatic mutation in the brains of germline heterozygotes. To further investigate this hypothesis, we established a conditional mouse model for *Depdc5*. Using CRISPR, we generated a floxed allele which, following the unilateral electroporation of *Cre* into developing brains, recombines to result in discrete regions of null tissue. Preliminary results show increased mTOR signalling and increased soma size in regions where the gene function has been abolished. Together, data from investigations using cell lines and mutant mice support the involvement of mTOR dysregulation in GATOR1-related epilepsy and a second-hit mechanism of disease.

137

FANCM protein maintains genome stability by removing co-transcriptional R-loops

Charlotte Hodson[1], Julienne J O'Rourke[1], Sylvie van Twest[1], Vincent J Murphy[1], Elyse Dunn[1], Andrew J Deans[1]

1. St Vincent's Institute, Fitzroy, VIC, Australia

R-loops are stable structures that form when RNA becomes trapped within duplex DNA. R-loops are important in regulation of cellular processes such as gene expression, telomere maintenance, immunoglobulin (Ig) class switching, and DNA repair. More recently, it has been shown that changes in R-loop regulation induce DNA damage and genome instability.

In particular, down-regulation of certain genes of the Fanconi Anaemia (FA) DNA repair pathway, results in an accumulation of R-loops. This suggests the FA pathway plays a role in R-loop metabolism and regulation [1, 2]. In particular, we have shown that the FA-associated gene FANCM can displace trapped RNA from an R-loop to maintain genome stability.

To further elucidate the mechanism by which the FANCM and the FA pathway regulates R-loop processing we have created a biochemical system for in vitro analysis of R-loops. We show that FANCM processively unwinds both short and long R-loops, irrespective of sequence, topology or coating by replication protein A. R-loops can also be unwound in the same assay by the yeast and bacterial orthologs of FANCM, Mph1 and RecG, indicating an evolutionary conserved function.

Consistent with this biochemical activity of FANCM, we show that FANCM deficient cells are sensitive to drugs that stabilize R-loop formation. **Our work reveals a mechanistic basis for R-loop metabolism that is critical for genome stability.**

1. 1. Schwab, et al., The Fanconi Anemia Pathway Maintains Genome Stability by Coordinating Replication and Transcription. Molecular Cell, 2015. 60(3): p. 351-361.
2. 2. García-Rubio et al., The Fanconi Anemia Pathway Protects Genome Integrity from R-loops. PLOS Genetics, 2015. 11(11): p. e1005674.

138

Hormone-mediated autophagy-dependent cell death in *Drosophila*

Donna Denton[1], Tianqi Xu[1], Sonia Dayan[1], Shannon Nicolson[1], Sharad Kumar[1]

1. Centre for Cancer Biology, University of South Australia, SA Pathology, Adelaide, South Australia, Australia

Autophagy and programmed cell death (PCD) are essential for animal development and for the maintenance of cell and tissue homeostasis. Autophagy is a conserved process that delivers components of the cytoplasm to lysosomes for degradation. Most PCD occurs by caspase-mediated apoptosis, with other mechanisms such as autophagy-dependent cell death, having crucial spatiotemporal restricted roles.

Drosophila is a powerful model to study hormone-regulated modes of programmed cell death (PCD) during development. During the late third instar larval stage, a large ecdysone pulse triggers the PCD of the obsolete larval tissues, including the larval midgut. In our seminal findings we discovered that autophagy acts as a cell death mechanism required for the removal of the larval midgut. Midgut PCD occurred normally in the absence of the apoptotic machinery, where as inhibition of autophagy blocked degradation. We found that hormonal cues together with growth arrest are prerequisite for autophagy-dependent midgut PCD.

As autophagy can act as either a cell survival or cell death mediator this raises the question of how autophagy specifically results in the demise of the cell. Our recent studies have shown that some components of the canonical autophagy pathway required for survival are dispensable for autophagy-dependent cell death. This suggests that autophagy-dependent cell death may involve distinct and novel regulatory proteins and mechanisms. We have been using proteomic and genetic approaches to identify tissue specific regulators of cell death verses cell survival autophagy. In further studies examining signalling pathways that regulate autophagy-dependent cell death we have now identified an unexpected role for the TGF-β homologue, Decapentaplegic (Dpp), as a key player in this process. The findings from these studies will be discussed.

139

Representing the human genome with chiral DNA reference standards

Ira Deveson[1], Bindu Swapna Kanakamedala[1], James Blackburn[1], Christopher Barker[1], Ted Wong[1], Andre Reis[1], Simon Hardwick[1], Timothy Mercer[1]

1. Garvan Institute of Medical Research, Sydney, NSW, Australia

An object that is inequivalent to its own mirror image is said to be *chiral*. Famously embodied in the human hand, this property also applies to nucleic acids. A DNA molecule of any length is inequivalent to a reverse-orientation copy of itself yet, just like the right and left hands, the two are similar in every respect. We have developed synthetic DNA standards (‘sequins’*)* that are simply reverse-orientation copies of natural DNA sequences. Thanks to their common properties, sequins behave similarly to the original sequences on which they were based, but can be easily and unambiguously distinguished from human DNA by alignment to a reverse-orientation reference. In this way, sequins can be used as internal controls during DNA sequencing, or other genetic assays, to provide qualitative and quantitative measures of analytic performance. This design principle can be used to synthetically represent virtually any feature of the genome, including natural instances of genetic variation, disease causing mutations or analytically challenging features, such as microsatellites or mobile elements. We have created a suite of >500 sequin standards, 1-10kb in length, that have been validated for use in whole genome and targeted next-generation sequencing, third-generation nanopore sequencing, and other assays. Sequins fulfill the unmet need for reference standards suitable for clinical genomics, an important step toward the realization of personalized medicine.

140

The Histone Variant H2A.Z is a Master Regulator of the Epithelial-Mesenchymal Transition

Renae Domaschenz[1], Sebastian Kurscheid[1], Maxim Nekrasov[1], Shuyi Han[1], David Tremethick[1]

1. John Curtin School of Medical Research, Australian National University, Canberra, ACT, Australia

Epithelial-mesenchymal transition (EMT) is a profound example of cell plasticity that is crucial for embryonic development and cancer. Although it has long been suspected that epigenetic-based mechanisms play a role in this process, no master epigenetic regulator that can specifically regulate EMT has been identified to date. Here, we show that H2A.Z can coordinate EMT by serving as either an activator or repressor of epithelial or mesenchymal gene expression, respectively. Specifically, TGFβ-β induces the loss of H2A.Z from the promoter regions of both epithelial and mesenchymal genes, but the functional outcomes of this loss are different; epithelial genes become repressed while mesenchymal genes are activated.

We demonstrate that H2A.Z nucleosomes positioned at different locations in a promoter can have different functions, i.e. the -2 nucleosome is involved in gene activation whereas the +1 nucleosome is linked to gene repression. This TGFββ-induced loss of H2A.Z from both epithelial and mesenchymal promoters is mimicked by the inhibition of H2A.Z expression. We suggest that H2A.Z has an important role in maintaining the epithelial state because its removal from certain promoters causes de-differentiation to the mesenchymal state. This is consistent with the finding that H2A.Z appears to regulate the expression of many more epithelial genes compared to the number of mesenchymal genes.

Most interestingly, one key gene that H2A.Z regulates is the *TGFB1β* gene itself. We found that depleting H2A.Z expression induced the TGFβ β autocrine loop suggesting that the ability of H2A.Z depletion to promote EMT is a combination of inducing the TGFβ β autocrine loop as well as directly regulating important EMT genes. This represents a previously unknown mechanism for the coordinated regulation of gene expression.

(Domaschenz et al., 2017, Cell Reports 21, 943–952)

141

VISUALISING THE NONSENSE: Fluorescent quantitation of the nonsense mediated mRNA decay pathway during brain development

Deepti Domingo[1], Jozef Gecz[1, 2, 3], David Bersten[1], Murray Whitelaw[1], Miles Wilkinson[4], Lachlan Jolly[1, 2]

1. The University of Adelaide, Adelaide, SA, Australia

2. Robinson Research Institute, Adelaide, SA, Australia

3. South Australian Health and Medical Research Institute, Adelaide, SA, Australia

4. The University of California, San Diego, California, America

The nonsense mediated mRNA decay pathway (NMD) is widely known as an RNA surveillance system, and more recently, as a global regulator of the transcriptome essential for embryonic viability. During brain development NMD plays an imperative role. Genetic variants in several key components of the NMD machinery which compromise NMD activity either cause, or are implicated in neurodevelopmental disorders. Evidence suggests that NMD activity is itself dynamically regulated, and shows variability at the cellular, tissue and whole organism level. Yet, current methods to quantify NMD activity are limited to biochemical end-point assays reporting on large populations of cultured cells, and as such, fail to capture the dynamic and heterogenous nature of the pathway at single cell level, in-vivo or in real time. To overcome these limitations, we have engineered a novel fluorescence based NMD reporter transgene. The transgene is comprised of three elements, namely the selection, control and NMD responsive expression cassettes. The control and NMD responsive cassette co-express distinguishable nuclear localised fluorescent proteins, with the ratio of expression (control: NMD responsive) providing visual and quantitative real-time readout of NMD activity. In addition, the transgene is conducive to standard protein and RNA based quantitative analysis. Using these methods we have shown our NMD reporter transgene to be responsive to NMD inhibition *in vitro*. The selection cassette utilises Flp/Frt recombination features to permit transgene insertion into the *Col1A1* locus of germ-line competent mouse ES cells, and thus the creation of NMD reporter mouse lines. This technology can provide real-time visual and quantitative tracking of endogenous NMD activity at a single cell resolution during embryonic development and into postnatal life, and will benefit our understanding of how NMD is regulated, and the biological processes in which it is employed, particularly relevant for our understanding of its role during brain development.

142

Genome-guided characterisation of evolved xylose metabolism in Saccharomyces cerevisiae

Richard J Edwards[1, 2], Asa Perez-Bercoff[1], Gustave Severin[1], Tonia L Russell[3], Psyche Arcenal[4], Anna Sophia Grobler[4], Paul V Attfield[4], Philip JL Bell[4]

1. School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia

2. NSW Systems Biology Initiative, University of New South Wales, Sydney, NSW, Australia

3. Ramaciotti Centre for Genomics, University of New South Wales, Sydney, NSW, Australia

4. Microbiogen Pty Ltd, Sydney, NSW, Australia

We have a unique opportunity to study the evolution of a new biochemical pathway in a sexually reproducing population. Microbiogen Pty Ltd have have successfully evolved a *Saccharomyces cerevisiae* (Baker's yeast) strain ("MBG11a") that, unlike wild strains, grows efficiently on xylose as a sole carbon source (Attfield & Bell 2006). Using PacBio long reads, we have sequenced and assembled MBG11a and the founding yeast strains from which the xylose metabolism evolved. We are using a combination of comparative genomics, deep

Illumina resequencing of the evolving population, RNA-Seq transcriptomics, and competition assays to identify genes involved in the evolution of this new metabolic activity.

Comparing our draft MBG11a assembly to the S288C yeast reference genome reveals over 115,000 variants, including over 14,000 coding changes across 3360 genes, plus has several large insertions and deletions (approx. 63 kb missing and 164 kb extra, excluding Ty element changes). To identify loci with important consequences for xylose utilisation, MBG11a was mated with a strain that could not grow on xylose or galactose. Haploid offspring were competed on glucose, galactose and xylose minimal media to identify MBG11a alleles increasing frequency under selection. Two candidate loci with unique protein-coding variants were identified in regions under significant positive selection on xylose minimal media. One showed classic signs of gene duplication followed by neofunctionalisation, while the other was identified as a master regulator of transcription. RNA-Seq analysis of MBG11a growing on xylose versus glucose highlighted two further candidate genes, previously shown to substitute for key xylose metabolic proteins.

Future work will focus on confirming whether the identified proteins are necessary and/or sufficient for *S. cerevisiae* to grow efficiently on xylose. Through "molecular palaeontology" on the evolving population, we will also trace how mutations have interacted with existing genetic variation to evolve this novel phenotype.

1. Attfield P.V. & Bell P.J.L. (2006). Use of population genetics to derive non-recombinant Saccharomyces cerevisiae strains that grow using xylose as a sole carbon source. FEMS yeast Res. 6:862-868.

143

MitoWisdom: An Unsupervised Mitochondrial Genome Analyser Using Deep Learning

Tansel Ersavas[1]

1. Garvan Institute of Medical Research, Darlinghurst, NSW, Australia

Mitochondria are critical to cell survival as the host cell's energy source and in regulating cell metabolism. Mitochondria's role in cancers, degenerative diseases and ageing are increasing in prominence, and better analysis tools are required to further identify their contributions to such conditions. We are developing a clustering mechanism that uses a novel deep learning system and unsupervised learning to extract features from mitochondrial genome data at multiple dimensions. This system then can be quickly and easily re-trained to analyse mitochondria in multiple ways with minimal sample data for specialised classification of any condition or trait. We use a "convolutional autoencoder" to reduce the dimensionality of the data and use the reducer part of the autoencoder as a basis of a trained DL system. We will then demonstrate that the generated encoder represents mitochondria and can be used as a knowledge source that can be applied to identify certain mitochondrial traits and conditions with minimal supervised training. The technique we use for the mitochondrial genome is general and is applicable to the whole genome or any selected proportions of it.

MitoWisdom currently uses sequenced DNA for its main input, but can also use RNA data as well. It extracts information from BAM or SAM files, then uses this digested ~16600x4 input to extract multiple-dimensional features from the DNA and/or RNA to shape the network. It is designed to be input agnostic, so as long the input can be converted to a NxM matrix it can be fed into the system. Once trained, the encoder part of the system can directly be used by itself or can be included as a part of a larger network that is trained to learn more complex inter-chromosome interactions.

144

Unsupervised demultiplexing of single-cell barcodes from raw nanopore sequencing data

James M Ferguson[1], Ghamdan A Al-Eryani[1], Manu Singh[1], Martin A Smith[1]

1. Garvan Institute of Medical Research, Darlinghurst, NSW, Australia

Single-cell RNAseq (scRNAseq) technology provides the ability to characterise individual cell types from heterogeneous samples, enabling gene expression profiling at deep resolution. In addition to limitations derived from short read sequencing, most scRNAseq technologies only resolve the 3' end of RNA transcripts, making it difficult or even impossible to characterise full molecules. Long-read sequencing platforms offer a solution to these problems, but come with a higher sequencing error-rate than Illumina sequencing and, consequently, make it difficult to accurately demultiplex cell barcodes and unique molecular identifiers (UMI) from scRNAseq data.

Here, we describe an unsupervised method to demultiplex full-length transcripts from single cells using Oxford nanopore sequencing. By vectorising the raw signal corresponding to the cell barcodes with a reference set using dynamic time warping, a growing neural gas can be used to map the topology in N-dimensional feature

space and cluster reads accordingly. A similar process for UMIs can generate a more accurate consensus sequence of the transcript, thus overcoming two significant technical limitations, and increasing the resolution and accuracy of scRNAseq.

145

Comparison of predicted and actual consequences of missense mutations

Matt Field[1], Lisa Miosge[2], Dan Andrews[2], Chris Goodnow[3]

1. James Cook University, Amithfield, QLD, Australia

2. Immunology, Australian National University, Canberra, ACT, Australia

3. Immunogenomics, Garvan Institute, Sydney, NSW, Australia

Each person's genome sequence contains thousands of missense variants. Practical interpretation of their functional significance must rely on computational inferences in the absence of exhaustive experimental measurements. Here we analyzed the efficacy of these inferences in 33 de novo missense mutations revealed by sequencing in first-generation progeny of ENU treated mice, involving 23 essential immune system genes. PolyPhen2, SIFT, MutationAssessor, Panther, CADD, and Condel were used to predict each mutation's functional importance, whereas the actual effect was measured by breeding and testing homozygotes for the expected in vivo loss-of-function phenotype. Only 20% of mutations predicted to be deleterious by PolyPhen2 (and 15% by CADD) showed a discernible phenotype in individual homozygotes. Half of all possible missense mutations in the same 23 immune genes were predicted to be deleterious, and most of these appear to become subject to purifying selection because few persist between separate mouse substrains, rodents, or primates. Because defects in immune genes could be phenotypically masked in vivo by compensation and environment, we compared inferences by the same tools with the in vitro phenotype of all 2,314 possible missense variants in tumour suppressor TP53; 42% of mutations predicted by PolyPhen2 to be deleterious (and 45% by CADD) had little measurable consequence for TP53-promoted transcription. We conclude that for de novo or low-frequency missense mutations found by genome sequencing, half those inferred as deleterious correspond to nearly neutral mutations that have little impact on the clinical phenotype of individual cases but will nevertheless become subject to purifying selection. These results highlight an important gap in our ability to relate genotype to phenotype in clinical genome sequencing: the inability to differentiate immediately clinically relevant mutations from nearly neutral mutations. This work was recently published in PNAS[1].

1. Miosge, L. A., M. A. Field, Y. Sontani, V. Cho, S. Johnson, A. Palkova, B. Balakishnan, R. Liang, Y. Zhang, S. Lyon, B. Beutler, B. Whittle, E. M. Bertram, A. Enders, C. C. Goodnow and T. D. Andrews (2015). "Comparison of predicted and actual consequences of missense mutations." Proc Natl Acad Sci U S A.

146

The Role of Twist in Glutamate Stimulated Neuronal Apoptosis

Elizabeth A Ford[1], Leslie Caron[1], Thang Khuong[1], John Manion[1], Greg Neely[1]

1. Charles Perkins Centre, University of Sydney, Camperdown, NSW, Australia

Posttraumatic glutamate release and subsequent excitotoxicity is a major contributor to neuropathic pain following nerve injury. The loss of GABAergic tone caused by excitotoxicity further imbalances the inhibition and excitation of neural circuits, leading to further neuronal apoptosis. It has been established that inhibiting caspase can stop loss of GABA neurons post injury, and that stopping this loss prevents heat allodynia. Drosophila Twist knockdown show dampened hypersensitivity in response to nerve injury. Twist is heavily implicated in embryonic development, however its role in apoptosis is relatively unknown. We now aim to show the effects of twist knockdown in human GABAergic neurons grown from human induced pluripotent stem cells. We have successfully grown GABAergic neurons in a 28 day cell culture protocol derived from human induced pluripotent stem cells. We now plan to create twist knockdown cell lines using siRNA, and test the effects of glutamate stimulation on these cells compared with control cells; we hypothesise the twist knock down cells will show reduction in apoptosis.

147

Analysis of ATAC sequencing data

Alexandra L Garnham[1, 2], Lisa A Mielke[1, 2, 3], Gabrielle T Belz[1, 2], Gordon K Smyth[1, 4]

1. The Walter and Eliza Hall Institute of Medical Research, Melbourne, Victoria, Australia

2. Department of Medical Biology, The University of Melbourne, Melbourne, Victoria, Australia

3. Olivia Newton-John Cancer Wellness & Research Centre, La Trobe University School of Cancer Medicine, Heidelberg, Victoria, Australia

4. Department of Mathematics and Statistics, The University of Melbourne, Melbourne, Victoria, Australia

The Assay for Transposase Accessible Chromatin with high-throughput sequencing or ATAC-seq, is a method for mapping genome-wide chromatin accessibility. Sequencing reads can be used to determine regions of increased accessibility, in addition to identifying locations of transcription factor binding and nucleosome position. Analysis of ATAC-seq data is most commonly performed using peak calling methods such as MACS2 or HOMER, which are ideal when the aim is to find accessible genomic regions. However, when performing a differential analysis between groups of samples, this approach may not be the most appropriate. It has been shown that calling peaks on individual samples or groups can cause a loss of error rate control during a differential analysis. Here we present two alternative approaches to ATAC-seq data analysis that overcome this difficulty. The first is a gene based approach where we focus on accessibility changes in the promoter region of all genes. This method is fast and easy to perform, as well as easily relatable to other sequencing technologies such as RNA-seq. The second approach applies the Bioconductor package CSAW which utilises sliding windows to analyse changes genome-wide. These methods are demonstrated and compared with the peak calling approach using an ATAC-seq data set that comprises mouse wild type and Tcf7 knock-out double negative 3 cells.

148

Understanding the role of cellular mosaicism in PCDH19 epilepsy and intellectual disability.

Jozef Gecz[1], Kristy Kolc[1], Claire Homan[1], Raman Sharma[1], Ingrid E Scheffer[2], Lynette Sadleir[3], Melissa Pitman[4], Slave Petrovski[2], Stuart Piston[4], Lachlan Jolly[1], Daniel Pederick[1], Paul Thomas[5], Duyen Pham[1]

1. Medicine, University of Adelaide, Adelaide, SOUTH AUSTRALIA, Australia

2. Medicine, The University of Melbourne, Melbourne, Victoria, Australia

3. Paediatrics, University of Otago, Otago, New Zealand

4. Centre for Cancer Biology, University of South Australia, Adelaide, South Australia, Australia

5. Medicine, University of Adelaide, Adelaide, South Australia, Australia

We have implicated procadherin 19 - *PCDH19,* in girls clustering epilepsy (GCE) in 2008. In excess of 10 males with somatic mosaic *PCDH19* mutation have also been reported. *PCDH19* evolved to be the second most important epilepsy gene. We have developed a multidimensional pathogenicity assessment tool for *PCDH19* using functional assays, *in silico* prediction and crystal structure modelling. We subsequently tested 25 variants; known disease-causing, VOUS as well as select, frequent population *PCDH19* variants. Our systematic review of 300 published cases shows that the penetrance of *PCDH19* GCE is about 70%, much less than >90% we estimated from family studies in 2008. The review also shows significant behavioural comorbidities (60%) and significant association between age at seizure onset and disease severity. No explanation for the variable penetrance between e.g., mothers and daughters or discordant MZ twins, have been identified. In this regard, we studied cellular and electrophysiological aspects of *PCDH19* GCE using CRISPR/Cas9 modified mouse models with both, wt and KO PCDH19 alleles visualised. Only the heterozygous females showed altered brain EEG activity (altered SWDs). These mice also showed highly specific sorting and distribution of PCDH19 wt and PCDH19 KO neurons in their developing cortices (14.5dpc). Using patient-derived and iPSC models we show that gene expression (OXTR, APOD, etc.) and behaviour of the wt and KO cells differ significantly (in e.g. cell polarity and differentiation). We postulate that individually specific X-inactivation together with altered gene expression & functionality of wt and mutant *PCDH19* cells are the underlying forces of PCDH19 GCE.

149

Mechanisms of regulation of the transcription factor Single-Minded 1 in obesity

Alexis G Gerassimou[1], Adrienne E Sullivan[1], David C Bersten[1], Murray L Whitelaw[1]

1. School of Biological Sciences, University of Adelaide, Adelaide, SOUTH AUSTRALIA, Australia

The basic helix-loop-helix/Per-Arnt-Sim (bHLH/PAS) transcription factor Single Minded 1 (SIM1) has essential roles in the development and function of hypothalamic cell lineages. Non-synonymous point mutants of *SIM1* have established SIM1 deficiency to be a monogenic cause of obesity in humans, while haploinsufficiency studies in mice revealed increased weight gain due to hyperphagia and increased linear growth. Several studies indicate that SIM1 plays an important role in the appetite controlling Leptin-Melanocortin signalling pathway in the hypothalamus, however the target genes of SIM1 and its up- and downstream regulatory pathways have yet to be defined. To further our understanding of the role of SIM1 in appetite control, we have generated a transgenic mouse model with GFP under the control of the *Sim1* promoter composite with a weakly functioning mutant of SIM1. This mouse model, in combination with cell based assays, will be used to identify novel target genes and investigate mechanisms of regulation of SIM1.

150

Expanding the repertoire of Transcription Factor Motifs

Kevin R Gillinder[1, 2], Graham W Magor[1, 2], Andrew C Perkins[1, 2]

1. Australian Centre for Blood Diseases, Monash University, Clayton, VIC, Australia

2. The Alfred Hospital , Melbourne, VIC, Australia

Site-specific transcription factors (TFs) play a critical role in the regulation of gene expression during mammalian development and disease. TFs recognise specific sequences of DNA, subsequently bind, and recruit transcriptional cofactors to regulate their target gene expression. However, genome-wide studies of TF occupancy, such as ChIP-seq, have revealed they bind relatively few of their cognate sequences. While, in addition, many sites that are bound *in vivo* lack identifiable binding sequences. This is likely due to cooperative binding from other transcriptional cofactors, or additional new binding sequences, or both.

To address this question, we have tested members of the Krüppel-like Factor (KLF), Specificity Protein (SP), and Early Growth Response (EGR) families – well-known members of the C2H2-ZF protein superfamily – for their ability to recognize previously unrecognised sequences; an expanded lexicon of binding sequences. Utilizing a combination of EMSA and next-generation sequencing – known as SPEC-seq – we have defined relative binding affinities and specificities for KLF, SP and EGR family members to more than a million variants of the 10bp binding *motif*.

We find a strong correlation between *in vivo* and *in vitro* relative binding affinities. Our results suggest TFs bind to motifs with a much broader spectrum of affinities than previously thought. This is likely because intermediate affinity sites are not adequately represented by SELEX motif enrichment approaches, phage display or by *de novo* motif discovery from ChIP-seq data, since these approaches are biased towards discovery of the very highest affinity sites. SPEC-seq is a new way to fully ascertain the full repertoire of potential binding sites for any TF, and thus provide new insights into the interactions between TF binding and epigenetic regulation of gene expression.

151

Modelling Breast Cancer Progression Using Single-cell RNA-seq

Brian S Gloss[1], Fatima Valdes-Mora[1], Robert Salomon[1], Yolanda Colino-Sanguino[1], Daniel Roden[1], Marcel Dinger[1], Christopher Ormandy[1], David Gallego-Ortega[1]

1. Garvan Institute, Darlinghurst, NSW, Australia

Cancer cell diversity constitutes a challenge for cancer treatment and deeply impacts the outcome of cancer patients. A simultaneous overview of cancer cells and associated stromal cells is critical for the design of improved therapeutic regimes. Single-cell RNA-seq has emerged as a powerful method to unravel heterogeneity of complex biological systems; this has enabled *in vivo* characterization of cell type compositions through unsupervised sampling and modelling of transcriptional states in single cells.

Here we use the cell type agnostic, high-throughput microfluidic-based, single-cell RNA-seq method *Drop-seq* to elucidate the function and cellular composition of breast tumours. We use the MMTV-PyMT ± Elf5 mouse mammary tumour model to provide high-resolution landscapes of the disease and highlight cellular events that result in the acquisition of the metastatic phenotype. We show breast cancer cell composition and

tumour heterogeneity with unprecedented definition, elucidating the cellular and molecular complexity of tumour progression within the context of a complex multicellular environment.

152

The Landscape Of Circular RNA Expression In The Human Brain

Akira Gokoolparsadh[1], Firoz Anwar[1], Irina Voineagu[1]

1. School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney

Circular RNAs are enriched in the mammalian brain and are upregulated in response to neuronal differentiation and depolarisation. These RNA molecules, formed by non-canonical back-splicing, have both regulatory and translational potential. Here we carried out an extensive characterisation of circRNA expression in the human brain, in nearly two-hundred human brain samples, from healthy individuals and autism cases. We identify thousands of novel circRNAs, characterise inter-individual variability of circRNA expression in the human brain, and identify brain-region specific circRNAs. We demonstrate that similarly to mRNAs, circRNA isoforms are not expressed stochastically, but rather as major isoforms, supporting their regulated expression in the human brain. We also find that circRNA expression is dynamic during brain development, decreasing with cellular maturation in brain organoids, but remains stable across the adult lifespan. These data provide a comprehensive catalogue of circRNAs and a deeper insight into their expression in the human brain.

153

Using long-read sequencing to detect haplotype-specific differential methylation

Scott Gigante[1], Alexis Lucattini[2], Andrew Keniry[1], Quentin Gouil[1], Matthew Tinning[2], Lavinia Gordon[2], Chris Woodruff[1], Matthew E Ritchie[1], Terry P Speed[1], Marnie E Blewitt[1]

1. Walter and Eliza Hall Institute, Parkville, VIC, Australia

2. Australian Genome Research Facility, Parkville, VIC, Australia

Asymmetric expression patterns between the two parental alleles are critical for development of the mammalian embryo. This process known as imprinting involves differential DNA methylation of the parental genomes. We sequence mouse embryonic placental tissue on the Oxford Nanopore MinION and exploit the long reads to determine both haplotype and CpG methylation levels. Comparison with matched Reduced-Representation Bisulfite Sequencing data confirms the accuracy of the methylation calls, and highlights the improvement in haplotyping conferred by the longer reads. We successfully identify known imprinting control regions, as well as novel differentially methylated regions. Based on their proximity to hitherto unknown monoallelically expressed genes, we propose that some of these regions could constitute new imprinting control regions.

154

The Human "Muscle-ome": Studying the skeletal muscle transcriptome to understand normal muscle biology and disease

Hayley Goullée[1, 2], Alistair Forrest[1, 2], Nigel Laing[1, 2], Rhonda Taylor[1, 2], Gina Ravenscroft[1, 2]

1. Harry Perkins Institute of Medical Research, Perth, WA, Australia

2. Centre for Medical Research, University of Western Australia, Perth, WA, Australia

Neuromuscular disorders (NMDs) are a group of >840 diseases affecting the peripheral nervous system and skeletal muscle. Despite increased identification of NMD disease genes, only half of patients receive a genetic diagnosis following testing. This is partially due to gaps in our knowledge on the coding and non-coding genes involved in muscle biology. Prioritizing potential muscle disease gene candidates first requires a comprehensive understanding of the genes that play a role in the healthy human “muscle-ome”.

Genetic variants tend to cause a phenotype in the tissues where the gene is highly expressed. We shall use CRISPR/Cas9 technology to delete genes that exhibit enriched expression in human skeletal muscle, but are not associated with disease. This will enable us to assess associated phenotypic and transcriptional changes, identify biological pathways that are affected, and infer the function of the target. We have selected candidate genes with enriched skeletal muscle expression within FANTOM5 data.

Initially targeting two long non-coding RNAs (lncRNAs) for deletion, we compared the resilience, transfection efficiency, and biological relevance of two myogenic cell lines: a clonal colony of immortalized Hu5/E18 myoblasts stably expressing Cas9-GFP, and primary human myoblasts transiently expressing Cas9-GFP. Results showed that the Hu5/E18s had significantly higher rates of survival, transfection and editing success than the primary myoblasts. However, the Hu5/E18s failed to differentiate into myotubes, making them less

biologically relevant than the primary myoblasts. We showed the primary myoblast lines have a normal karyotype, whereas both the clonal Hu5/E18 populations and the originally supplied Hu5/E18 cells have abnormal ploidy.

Future directions include optimizing the methods of transfection, exploring the utility of single cell transcriptomic profiling and editing of iPSCs, which can then be differentiated down the muscle lineage.

155

atz-1 promotes meiosis to maintain germline chromosomal integrity

Taylin E Gourley[1], Joseph A Dawson[1], Caitlin Methven-Kelley[1], Gregory M Davis[1]

1. Federation University Australia, Hazelwood South, VIC, Australia

Sexually reproducing species rely on the exchange of genetic information between homologous chromosomes in a process referred to as meiosis. This can be summarised as one round of DNA replication accompanied by two rounds of chromosome segregation to produce haploid gametes from diploid cells. This process involves tight coordination of a meiotic specific cohesion complex, the synaptonemal complex, and DNA damage repair mechanisms. We have investigated a gene encoding a putative myosin heavy chain protein, which we have named *atz-1* (Abnormal Transition Zone) as a novel gene that is required for maintaining germline chromosome integrity. Deletion of *atz-1* results in reduced brood size and marked embryonic lethality. Interestingly, *atz-1* mutants display a depleted, or absent transition zone, accompanied by reduced expression of the meiotic cohesion protein, REC-8. *atz-1* mutants display downstream germline defects including an extended pachytene region, sperm defect, aggregated chromosomal bodies, endomitotic oocytes, and elevated germ cell apoptosis. In addition to this, *atz-1* mutants are also hypersensitive to mild inhibition of DNA damage repair by hydroxyurea, suggesting that the initial round of DNA replication in *atz-1* mutants is impaired. Moreover, the *atz-1* mutant phenotype is germline specific and resupplying somatically expressed *atz-1* does not rescue the reproductive defects associated with *atz-1* mutants. Overall, our data suggests that *atz-1* is required for promoting meiosis to maintain germline chromosomal integrity.

156

Development of Molecular Therapies for *PRPF31*-associated Retinitis Pigmentosa Using Splice-switching Antisense Oligonucleotides

Janya Grainok[4, 2, 3, 1], Ianthe Pitout[2, 1], May T. Aung-Htut[2, 1], Fred K. Chen[5, 6, 7], Steve Wilton[2, 1], Sue Fletcher[2, 1]

1. Centre for Comparative Genomics, Murdoch University, Western Australia, Australia

2. Perron Institute for Neurological and Translational Science, Nedlands, Western Australia, Australia

3. Department of Biochemistry, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

4. Murdoch University, Murdoch, Western Australia, Australia

5. Centre for Ophthalmology and Visual Science, , University of Western Australia, Western Australia, Australia

6. Lions Eye Institute, Nedlands, Western Australia, Australia

7. Department of Ophthalmology, Royal Perth Hospital, Western Australia, Australia

Background: Pre-mRNA processing factor 31 (*PRPF31*) is essential for pre-mRNA splicing. Heterozygous loss of function mutations in *PRPF31* cause autosomal dominant retinitis pigmentosa 11 (RP11), characterized by photoreceptor degeneration, with potential for blindness. Increasing the *PRPF31* transcript level in the retina may rescue splicing function and therefore prevent or delay degeneration of photoreceptor cells. This study aims to upregulate *PRPF31* expression, sufficient to support retinal function in an RP11 patient carrying c.1205 C>A nonsense mutation in exon 12 of *PRPF31*. Splice-switching antisense oligonucleotides (AOs) were designed to exclude exon 12 and restore the reading frame of the *PRPF31* transcript, facilitating the production of potentially functional PRPF31 in RP11 patient cells.

Methods: AOs targeting splicing enhancer motifs in *PRPF31* exon 12 were designed to promote exclusion of exon 12 during pre-mRNA processing, and transfected into RP11 fibroblasts. RNA was isolated 24 hours after transfection and assessed for *PRPF31* exon 12 exclusion using end-point RT-PCR and total *PRPF31* transcript using quantitative RT-PCR.

Results: An AO targeting splicing enhancer sequences at the position 70-94 on exon 12 excluded exon 12 from 80% of the *PRPF31* transcripts, as determined by RT-PCR and Sanger sequencing. Exclusion of exon 12 resulted in approximately two-fold increase in total *PRPF31* transcript at the mRNA level, compared to control.

Conclusion: Exclusion of the exon bearing the nonsense mutation restored the reading frame of *PRPF31* and increased total *PRPF31* transcript, which would otherwise have been degraded through nonsense mediated

decay. Deletion of exon 12 does not disrupt any predicted functional domain and therefore the truncated protein is expected to retain some function. PRPF31 protein expression and the consequences on the photoreceptors will be further evaluated in retinal organoids derived from patient induced-pluripotent stem cells. This study will yield a potential candidate for *PRPF31*-associated RP.

157

Investigating the role of histone acetylation in angiogenesis

Zoe L Grant[1, 2], Tim Thomas[1, 2], Joan K Heath[1, 2], Anne K Voss[1, 2], Leigh Coultas[1, 2]

1. Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia

2. Department of Medical Biology, University of Melbourne, Melbourne, Victoria, Australia

Publish consent withheld

158

Improved Chemistry for NGS Library Cleanup and Size Selection

David Grasso[1], Charles Cowles[2], Douglas Horejsh[2], Mark Denhart[2], Curtis Knox[2], Adam Blatter[2], Marjeta Urh[2]

1. Promega Corporation, Auburn, Victoria, Australia

2. Promega Corporation, Madison, Wisconsin, USA

Next Generation Sequencing (NGS) libraries require high quality nucleic acid inputs of varying quantities, concentration, and size depending on the library preparation methods and sequencing platforms used. Regardless of these variations, in most instances a magnetic bead-based chemistry is utilised as a portion of the overall protocol. The steps using magnetic bead chemistries fall into two basic categories of function:

1) Sample cleanup: Removes sequencing adaptors or PCR primers, dNTP's, enzymes or unwanted buffer formulations.

2) Size Selection: Removes unwanted nucleic acid fragment or library molecules that are above or below a specified size range optimal for the downstream sequencing platform.

By varying the ratio of bead chemistry added to the original volume of DNA in solution, the user can alter the size of DNA captured by the beads or left behind in solution.

159

Optimisation of massively parallel single cell RNA-seq method (Drop-seq) in tumour tissues

Kristina Handler, Rob Salomon, Brian Gloss, Lesley Castillo, Andrew Law, Daniel Roden, Marcel Dinger, Christopher Ormandy, Fatima Valdes Mora, David Gallego Ortega

In the last few years single cell-transcriptome analysis has revolutionized our understanding of biology. Unlike bulk RNA sequencing, it enables the precise definition of cell populations within complex tissues, enabling the discovery of new cell types and cell states.

One of the major challenges for single-cell transcriptomics is the high cost and low throughput of RNA-seq per cell. The recent introduction of cell barcoding and microfluidics has dramatically decreased the cost per cell. Drop-seq is the most comprehensive and low-cost microfluidic-based method. It encapsulates individual cells into nanoliter-sized aqueous droplets, associating a different barcode with each cell's RNAs, and sequencing them all together.

Currently Drop-seq has not been implemented in tumour tissues and its optimization in this system would provide detailed information of the multiple cell populations that coexist in a tumour, such as cancer epithelial cells, immune cells and stromal cells. Therefore, for the first time, we have optimised the Drop-seq method in breast cancer using the MMTV-PyMT mouse mammary tumour model. Here we present the standard quality checks to demonstrate that our method is robust and comparable to any other single cell RNAseq method in other well-established tissues. Our quality controls include: mixing species experiments, the analysis of the presence of mitochondrial genes, batch effects and the cell clusters generation in relation to the number of cells analysed. Technical noise has been modelled comparing fresh tumour samples before and after their incubation at 4C degrees for 24h. This last quality check is important to apply this for patient samples where a precise protocol for sample preparation would be needed.

We have now established Drop-seq in tumour tissues. This new protocol has a huge potential for the high-throughput analysis of single cells in tumours at a low cost that could be easily implemented in any lab or hospital.

160

Transcriptional complexity of non-coding genomic regions associated with cognitive function

Simon A Hardwick[1, 2], Wendy Y Chen[1], Ted Wong[1], Ira W Deveson[1], James Blackburn[1], Timothy R Mercer[1, 3, 2]

1. Genomics & Epigenetics Division, Garvan Institute of Medical Research, Sydney, NSW, Australia

2. Faculty of Medicine, UNSW Sydney, Sydney, NSW, Australia

3. Altius Institute for Biomedical Sciences, Seattle, WA, United States

RNA sequencing (RNA-seq) can be used to assemble spliced isoforms, quantify expressed genes and provide a global profile of the transcriptome. However, the size and diversity of the transcriptome, the wide dynamic range in gene expression and inherent technical biases confound RNA-seq analysis. We have developed a set of spike-in RNA standards, termed 'sequins' (sequencing spike-ins), that represent full-length spliced mRNA isoforms. Sequins have an entirely artificial sequence with no homology to natural reference genomes, but they align to gene loci encoded on an artificial *in silico* chromosome. The combination of multiple sequins across a range of concentrations emulates alternative splicing and differential gene expression, and it provides scaling factors for normalization between samples. We demonstrate the use of sequins in RNA-seq experiments to measure sample-specific biases and determine the limits of reliable transcript assembly and quantification in accompanying human RNA samples. In addition, we have designed a complementary set of sequins that represent fusion genes arising from rearrangements of the *in silico* chromosome to aid in cancer diagnosis. RNA sequins provide a qualitative and quantitative reference with which to navigate the complexity of the human transcriptome.

161

RNA splicing in mammalian sex determination

Allara Zylberberg[1], Aleisha Symon[1], Daniel Bird[1], Stefan Bagheri-Fam[1], Francis Poulat[2], Vincent Harley[1]

1. Hudson Institute of Medical Research, Melbourne, VIC, Australia

2. Institut de Génétique Humaine, Montpellier, France

In mammals, male fate is under the control of the master transcriptional regulator, SOX9: in its presence, somatic precursor cells of the embryonic gonads differentiate into Sertoli cells, the central organizers of testicular differentiation. Therefore, analyzing target genes of this transcription factor allows the elucidation of cellular commitment mechanisms at the genome level. ChIP-seq in murine and bovine wild-type testes combined with RNAseq from mouse testes lacking SOX9 was undertaken[1]. 142 sex-specific RNA splicing events were detected in the XY knockout testes, similar to XX ovaries. Surprisingly, nearly half these genes were were bound by SOX9 in wild type testes. This suggests that SOX9 mediates both transcription and differential splicing of its target genes, as seen in chondrogenesis[2]. FGFR2 has two splice isoforms and the 2c form is essential during but not after sex determination[3]. Validation of RNAseq/Chipseq data in a human Sertoli cell line indicated that SOX9 mediates differential RNA splicing of the FGFR2 gene through DNA binding. Our results suggests that RNA splicing of FGFR2 could be a fine-tuning mechanism to reduce FGF9 ligand binding function once sex has been determined and that RNA splicing could be important in sex determination in mammals, as in Drosophila.

[1]Lavery et al., NAR 2017

[2]Hata et al., JCI 2011

[3]Bagheri-Fam et al., Endocrinology 2017

162

SRY and male sex bias in Parkinson's disease

Joo Lee[1, 2], Paulo Pinares-Garcia[1, 2], Hannah Loke[1], Seungmin Ham[1], Eric Vilain[3], Vincent Harley[1, 2]

1. Hudson Institute of Medical Research, Melbourne, VIC, Australia

2. Department of Anatomy and Developmental Biology, Monash University, Melbourne, Victoria, Australia

3. Human Genetics, University of California , Los Angeles, California, USA

Parkinson's disease (PD) results from the selective loss of dopaminergic neurons from the substantia nigra compacta (SNc). Whilst the cause of dopamine cell loss in Parkinson's disease (PD) is unknown, it is clear that the male-sex is a strong risk factor. The incidence and prevalence of PD is 2-fold higher and disease progression more rapid in males than females. Growing evidence suggests that sex-specific genes contribute to this male-bias in PD. We previously showed that the male-sex determining gene encoded by the Y

chromosome and therefore only found in males, SRY, co-localises with male dopamine neurons, where it regulates dopamine biosynthesis and motor function.Here, we investigated the regulation and function of nigral SRY in normal and 6-hydroxydopamine (6-OHDA) or rotenone lesioned hemiparkinsonian rats. Results show that *SRY* has a detrimental effect in males in animal models of Parkinson's disease. Specifically, we found that mid-brain *SRY* levels are aberrantly upregulated. Infusing antisense *SRY* DNA to downregulate *SRY* levels in the mid-brain led to a reduced in movement deficits and brain cell death. We also observed marked reductions in DNA damage, mitochondrial degradation and neuroinflammation in the mid-brain. These results suggest contribution to male sex bias in Parkinson's disease with implications for ADHD, autism and schizophrenia.

163

Confident effect sizes controlling FDR provide an ideal ranking of differentially expressed genes

Paul F Harrison[1], Andrew D Pattison[2], David R Powell[1], Traude H Beilharz[2]

1. Monash Bioinformatics Platform, Monash University, Clayton, VIC, Australia

2. Biomedicine Discovery Institute, Monash University, Clayton, VIC, Australia

A method is described for giving a confidence bound on the magnitude of Log Fold Change (LFC) in gene expression while controlling the False Discovery Rate (FDR). We propose this confidence bound as an ideal quantity by which to rank differentially expressed genes when presenting the results of an RNA-seq experiment. The method builds on the TREAT method of McCarthy and Smyth (2009). Unlike TREAT, a minimum LFC of interest does not need to be specified. The only parameter is the desired FDR, for which a reasonable default value can be given.

Sorting by p-value is a common default in the output of differential expression software. We compare this to our method of ranking genes, using a breast cancer RNA-seq data-set consisting of matched tumor-normal pairs. The top genes as ranked by p-value have small but consistent differential expression, whereas the top genes as ranked by confidence bound have a much larger magnitute of differential expression but also higher variability. This leads to a difference in biological interpretation, with greater emphasis placed on genes related to the extra-cellular matrix by our confidence bound method.

The confidence bound method degrades gracefully on subsets of samples in this data-set. For experiments with low statistical power, the ranking is similar to the p-value ranking, but as the power of an experiment increases the ranking is increasingly determined by the true effect size. Comparing the confidence bound and estimated LFC of top genes provides immediate feedback on whether or not an experiment was under-powered. As such, we propose our method as a better default method of ranking differentially expressed genes.

An R package implementing the method is available at https://github.com/pfh/topconfects

1. McCarthy, D. J., and Smyth, G. K. (2009). Testing significance relative to a fold-change threshold is a TREAT. Bioinformatics 25, 765-771. http://bioinformatics.oxfordjournals.org/content/25/6/765

164

A highly selective mechanism to discriminate transposon RNA from self-RNA in *Drosophila* ovaries

Rippei Hayashi[1], Dominik Handler[2], Julius Brennecke[2]

1. The Australian National University, Acton, ACT, Australia

2. Institute of Molecular Biotechnology of Austrian Academy of Sciences (IMBA), Vienna, Austria

Transposons are parasitic genetic elements that are present in all eukaryotic genomes sequenced to date. Uncontrolled expression of transposons threats the genome integrity, therefore the expression of transposons is tightly suppressed by host defence mechanisms. piRNA-guided gene silencing plays a central role in silencing transposons in animal gonads. Unlike micro-RNAs and siRNAs, piRNAs are made from single-stranded RNA by a Dicer-independent mechanism. piRNA biogenesis is highly selective for transposon anti-sense RNA, yet is thought to have a flexibility and adaptivity in order to deal with newly invaded transposon sequences. No sequence motifs or secondary structures in RNA have been shown to guide the production of piRNAs.

In the present work, we report that the RNA helicase fs(1)Yb selects substrate RNAs for piRNA production in *Drosophila* ovaries. Upon depletion of fs(1)Yb, piRNAs derived from the genomic locus called *flamenco*, which produces around a half of transposon anti-sense piRNAs, are severely reduced while other genomic loci including mRNAs start to produce more piRNAs compared to wild type cells. We will discuss about our model of how fs(1)Yb selects flamenco RNA against other cellular RNAs.

165

Protein recoding by ADAR1-mediated RNA editing is not essential for normal development and homeostasis

Jacki E Heraud-Farlow[1], Alistair Chalk[1], Sandra Linder[2], Qin Li[2], Scott Taylor[1], Ankita Goradia[1], Brian Liddicoat[1], Jin Billy Li[2], Carl Walkley[1]

1. St Vincent's Institute, Fitzroy, VIC, Australia

2. Department of Genetics, Stanford University, Stanford, California, USA

Adenosine-to-inosine (A-to-I) editing of dsRNA by ADAR proteins is a pervasive epitranscriptome feature. Tens of thousands of A-to-I editing events are defined in the mouse, yet the functional impact of most is unknown. Editing causing protein recoding is the essential function of ADAR2, but an essential role for recoding by ADAR1 has not been demonstrated. ADAR1 has been proposed to have editing-dependent and editing-independent functions however, the relative contribution of these *in vivo* has not been clearly defined. A critical function of ADAR1 is editing of endogenous RNA to prevent activation of the dsRNA sensor MDA5 (*Ifih1*). Concurrent deletion of MDA5 rescues the embryonic lethality Adar1 editing deficient mice (*Adar1*$^{E861A/E861A}$*Ifih1*). Outside of this role, it is uncertain how ADAR1 editing contributes to normal development and homeostasis.

We describe the consequences of ADAR1 editing deficiency on murine homeostasis in the absence of MDA5. *Adar1*$^{E861A/E861A}$*Ifih1*$^{-/-}$ mice are strikingly normal, including their lifespan. There is a mild, non-pathogenic innate immune activation signature in the *Adar1*$^{E861A/E861A}$*Ifih1*$^{-/-}$ mice. Assessing A-to-I editing across adult tissues demonstrates that outside of the brain, ADAR1 performs the majority of editing and that ADAR2 cannot compensate in its absence. Direct comparison of the *Adar1*$^{-/-}$and *Adar1*$^{E861A/E861A}$ alleles demonstrates a high degree of concordance on both *Ifih1*$^{+/+}$ and *Ifih1*$^{-/-}$ backgrounds, suggesting no substantial contribution from ADAR1 editing-independent functions under homeostatic conditions. These analyses demonstrate that the lifetime absence of ADAR1-editing is well tolerated in the absence of MDA5. We conclude that protein recoding arising from ADAR1-mediated editing is not essential for organismal homeostasis. Additionally, the phenotypes associated with loss of ADAR1 are the result of RNA editing and MDA5-dependent functions.

166

Transcriptional Consequences of Cancer Fusion Genes

Erin E Heyer[1], Ira W Deveson[1, 2], Timothy R Mercer[1, 3, 4], James Blackburn[1, 3]

1. Garvan Institute of Medical Research, Darlinghurst, NSW, Australia

2. School of Biotechnology and Biomolecular Sciences, Faculty of Science, University of New South Wales, Randwick, NSW, Australia

3. St. Vincent's Clinical School, Faculty of Medicine, University of New South Wales, Randwick, NSW, Australia

4. Altius Institute for Biomedical Sciences, Seattle, WA, USA

Chromosomal translocations join together two previously distinct parts of the genome. This genomic shuffling can alter expression levels of canonical genes or join two separate genes to encode a single transcript of novel function. These fusion genes account for approximately 20% of human cancer morbidity and are often tumour-specific and drug-targetable, so precise fusion identification can significantly influence disease treatment. To that end, we have developed Blood FuSeq and Solid FuSeq – two diagnostic tests utilising targeted RNA capture sequencing (RNA CaptureSeq) to identify fusion genes in haematological malignancies and solid tumours, respectively.

During initial validation using cell lines harbouring fusion genes, we precisely identified all known fusions and established the sensitivity of the assays at 1:10,000 cells. Expanding our analysis to patient samples, we successfully diagnosed both known and previously overlooked fusion genes, including those for which approved therapeutic treatments already exist. For example, we used the Blood FuSeq panel to interrogate 29 patient samples previously lacking a molecular diagnosis, and in doing so identified fusion genes in 16 of these patients.

Beyond identifying the genes involved in each fusion rearrangement, our data revealed informative variations in exon usage, expression levels and isoform diversity – all factors that can affect treatment efficacy. In some patients, we detected either multiple fusion gene isoforms or novel intragenic deletions undetected by other diagnostic approaches. Given the high coverage data generated by RNA CaptureSeq, we also expanded our analysis to search for novel transcriptomic elements unique to these cancer genomes.

Overall, we believe these FuSeq panels can be developed into clinical diagnostic tools, delivering medically relevant information on a wide range of fusion genes. In addition, the resulting expression level and exon

usage information for general cancer-associated genes can be used to infer further therapeutic treatments, all within a single diagnostic test.

167

Therapeutic potential of antisense oligonucleotide-mediated exon inclusion for Stargardt disease

Di Huang[2, 3, 1], Sue Fletcher[3, 1], May Aung-Htut[3, 1], Norman Palmer[3], Fred Chen[2, 4], Steve Wilton[3, 1]

1. Molecular Therapy Laboratory, Murdoch University, Perth, Western Australia, Australia

2. Centre for Ophthalmology and Visual Science & Lions Eye Institute,The University of Western Australia, Perth, Western Australia, Australia

3. Perron Institute for Neurological and Translational Science, Perth, Western Australia, Australia

4. Department of Ophthalmology, Royal Perth Hospital, Perth, Western Australia, Australia

Stargardt disease (STGD1) is an autosomal recessive juvenile-onset macular dystrophy caused by mutations in the ATP-binding cassette transporter gene (*ABCA4*). ABCA4 is localized to the rims of the outer segments disc membrane of photoreceptors and plays a crucial role in the removal of excessive 11-*cis* and all-*trans* from discs. Mutations that cause loss of function/mislocalization/misfolding of ABCA4 are thought to lead to a progressive accumulation of cytotoxic bisretinoid compounds and eventually to death of photoreceptors and retinal pigment epithelium cells. Currently, the consequences of most mutations are not known and there is no effective treatment.

The c.5461-10T>C mutation in *ABCA4* gene was reported to influence *ABCA4* transcript processing, resulting in mature transcripts missing exon 39, and exon 39 and 40. Deletion of either or both exons causes a frameshift in the *ABCA4* transcript that likely leads to nonsense mediated decay. Splice modulating antisense oligonucleotides (AOs) have been shown to be effective in exon inclusion in other conditions, and we hypothesized that the same strategy may be applicable to STGD1 patients with the c.5461-10T>C mutation.

We designed several AOs targeting intronic splicing silencers within intron 39 to enhance inclusion of exon 39 or exon 39 and 40, in the mature *ABCA4* transcript. The AOs were transfected into patient fibroblasts carrying the c.5461-10T>C mutation and the *ABCA4* transcript was analyzed after 24 hours. An increase in the full-length *ABCA4* transcript was observed in treated compared to the untreated cells. This increase is not only through exon 39 inclusion but also prevention of nonsense mediated decay. A decrease in the transcript missing exon 39 was also observed. Further studies to assess ABCA4 protein expression and function after treatment are in progress.

This study demonstrates that a splice modulating AO has therapeutic potential for STGD1 patients carrying the c.5461-10T>C mutation through splice correction.

1. Sangermano R, Bax NM, Bauwens MPhotoreceptor et al. Progenitor mRNA Analysis Reveals Exon Skipping Resulting from the ABCA4 c.5461-10T→C Mutation in Stargardt Disease. Ophthalmology. 2016 Jun;123(6):1375-85.
2. Molday RS. Insights into the Molecular Properties of ABCA4 and Its Role in the Visual Cycle and Stargardt Disease. Prog Mol Biol Transl Sci. 2015;134:415-31.
3. Aukrust I, Jansson RW, Bredrup C et al. The intronic ABCA4 c.5461-10T>C variant, frequently seen in patients with Stargardt disease, causes splice defects and reduced ABCA4 protein level. Acta Ophthalmol. 2017 May;95(3):240-246.

168

Genome-wide discovery of translation control mechanisms

Yoshika Janapala[1], Nikolay Shirokikh[1], Traude Beilharz[2], Stuart Archer[2], Thomas Preiss[1]

1. John Curtin School of Medical Research,ANU,Acton, 131 Garran Rd, Acton, ACT, Australia

2. Monash University, Melbourne, Victoria, Australia

The control of mRNA translation into proteins is critical for the adaptation of eukaryotic cells to environmental changes and stress conditions. Nutrient stress responses in yeast (*Saccharomyces cerevisiae*) are prototypical to eukaryotes and are known to cause large-scale translational reprogramming, downregulating translation of most messenger (m)RNAs while upregulating translation of others. Glucose starvation is one of the better studied nutrient stresses in yeast as it causes rapid changes, mainly during the initiation phase of translation.

While this effect has been described before, methods to obtain insights into particular initiation intermediates, and the dynamics of the response, were limiting. To understand how translational control operates, we aim to provide detailed data for the response to glucose starvation, a common stress condition where dysregulation can be critical in multiple diseases including cancer.

Glucose starvation induces rapid loss of the translation initiation factor eIF4A from initiation complexes, followed by relocation of these stalled initiation intermediates into cytoplasmic bodies. However, the mRNA-wide loss of eIF4A, its dynamics upon the induction of the stress, and its specific effects on translation across the transcriptome remain unknown.

To address these problems, we refined TCP-seq, which is based on ribosome footprinting (Archer SK et al. Nature 2016 535:570-574; Shirokikh NE et al. Nat. Protoc. 2017 12:697-731). TCP-seq captures all ribosome-mRNA complexes, including initiation intermediates. In glucose starvation/restoration time-course experiments, we found that the translation response time, measured by changes in polysome profiles, was much shorter than anticipated, approaching 20 seconds. We selected three conditions that resulted in minimal, intermediate or maximal polysome disassembly, to obtain samples for TCP-seq. Further, we aim to use yeast strains harbouring tagged versions of eIF4A, eIF4E, eIF4G and Pab1 for selective footprinting of complexes involving these factors and uncover their mRNA-specific distribution and its dynamics across the transcriptome during glucose starvation.

169

The role of Smchd1 in mediating higher order chromatin conformation

Natasha Jansz[1], Andrew Keniry[1], Kelan Chen[1], Kelsey Breslin[1], Tamara Beck[1], James Murphy[1], Marnie E Blewitt[1]

1. Walter and Eliza Hall Institute, Parkville, VIC, Australia

Structural Maintenance of Chromosomes, Hinge Domain containing 1 (Smchd1) is critical for the maintenance of X Chromosome Inactivation (XCI), and transcriptional repression at a subset of autosomal loci. Gain and loss of function mutations in *SMCHD1* underlie Bosma arhinia micropthalmia syndrome (BAMS) and Facioscapulohumoral muscular dystrophy 2 (FSHD2), respectively - two distinct developmental disorders. Currently little is known about molecular mechanisms underlying the involvement of Smchd1 in transcriptional repression or disease.

Evidence in the literature suggests that Smchd1 and the non-coding RNA *Xist* might interact directly. We have previously shown that the hinge domain of Smchd1 binds synthetic oligonucleotides in vitro. I was therefore interested in whether Smchd1 directly associates with endogenous nucleic acids, and whether such interactions could be important for Smchd1's localisation to the chromatin. To this end, I performed PAR-CLIP to determine whether Smchd1 binds endogenous RNAs genome-wide. I find Smchd1-RNA interactions to be non-specific, and are therefore unlikely to act as a targeting mechanism. I also find that while Smchd1 is dependent on *Xist* for its localisation to the Xi, this is not due to a direct protein-RNA interaction, but rather due to a dependency on the downstream HnrnpK-polycomb pathway.

There is evidence to suggest that Smchd1 may be involved in regulating higher order chromatin organisation. To investigate changes to the chromatin architecture in the absence of Smchd1, I performed in-situ Hi-C and ATAC-seq in *Smchd1* wild-type and deleted neural stem cells. For the first time my data have demonstrated a role for Smchd1 in chromatin organisation of the *Hox* clusters, and the inactive X chromosome. Furthermore, I have identified that in the absence of Smchd1, *Hox* genes are dysregulated, implicating Smchd1 in *Hox* gene silencing via a role in chromatin conformation.

170

Whole genome sequencing is improving the identification of genetic causes of paediatric cataracts, including novel cataract genes.

Johanna L Jones[1], Bennet McComish[1], Mark A Corbett[2], Jac Charlesworth[1], Jamie E Craig[3], Kathryn P Burdon[1]

1. Menzies Institute for Medical Research, University of Tasmania, Hobart, Tasmania, Australia

2. School of Medicine, University of Adelaide, Adelaide , South Australia, Australia

3. Department of Ophthalmology, Flinders University, Bedford Park, South Australia, Australia

Whole genome sequencing is enhancing the ability to identify disease causing mutations in inherited paediatric cataracts (PC), a heterogeneous rare disease that causes visual impairment. We are investigating unsolved families from Australia's largest repository of PC DNA, which contains 191 families. Nine large families, with at least 3 affected individuals each, have been selected for linkage analysis. All available DNA samples (n=96)

have undergone genome-wide genotyping using Illumina OmniExpress SNP arrays. Parametric linkage analysis, using Merlin, is being performed with a relevant disease model for each family. Whole genome sequencing of an affected individual from each family was performed on an Illumina X10 platform. The Churchill pipeline was used to align fastq files to hg19, and variant calling was performed using SAMtools for SNPs and Platypus for indels. Structural variants will also be considered. Variants within putative linkage regions are filtered for frequency (MAF <0.01) and predicted functionality (using tools such as CADD). Using this pipeline, the PC in family CRVEEH66 have been linked to a 6.8Mb region at Xq24 (LOD=2.53) and two additional regions at 1q42.2-1q43 and 3q26.31-3q26.32 (LOD=2.44). A Complete Genomics™ whole genome sequence of the proband enabled the identification of a 127kb deletion that truncates the *PGRMC1* gene following exon 1 and completely removes a long non-coding RNA gene *LOC101928336,* located in the chromosome Xq24 region. This variant was not detectable by prior exome or Sanger sequencing and these are novel PC candidate genes. This highlights the utility of whole genome sequencing for identifying putative disease causing structural and non-coding variants, which have not been explored in PC. This work will provide these Australian families with a molecular diagnosis and enable genetic screening. The discovery of novel genes will expand the suite of genes available for screening in other patients and improve our understanding of cataract pathogenesis.

171

Loss of topoisomerase III beta is associated with hallmarks of genome instability

Paul Kalitsis[1, 2], Tao Zhang[1], Mathew Wallis[3], Damien F Hudson[1, 2]

1. Murdoch Childrens Research Institute, Parkville, VICTORIA, Australia

2. Department of Paediatrics, University of Melbourne, Parkville, Victoria, Australia

3. Clinical Genetics Services, Austin Health, Heidelberg, Melbourne, Victoria, Australia

Topoisomerase III beta (TOP3B) is a conserved type IA nucleic acid processing enzyme that catalyses RNA strand passage reactions. Together with partner proteins, TOP3B is implicated in the resolution of R-loops that are generated during transcription. Previous studies in knockout mice show a slight reduction in lifespan, chromosomal defects in germ cells and progressive infertility across generations. Furthermore, mouse embryonic fibroblasts are sensitive to DNA-damaging agents and have impaired up-regulation of p53. To further understand whether the loss of TOP3B increases genome instability events we have generated knockout human cell lines. We show elevated rates of genome instability in the form of micronuclei, gH2Ax signals and chromosome segregation defects. We are additionally investigating a family with a homozygous deletion of the TOP3B gene with overlapping phenotypes observed in cell line and animal model systems.

172

Metabolic and transcriptomic analyses of Müller glial cells of the retina

Thaksaon Kittipassorn[1, 2], Cameron D. Haydinger[1], John P. Wood[3, 4], Teresa Mammone[3, 4], Robert J. Casson[3, 4], Daniel J. Peet[1]

1. Department of Molecular & Cellular Biology, School of Biological Sciences, University of Adelaide, Adelaide, SA, Australia

2. Department of Physiology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

3. Ophthalmic Research Laboratories, Hanson Institute, Adelaide, SA, Australia

4. South Australian Institute of Ophthalmology, University of Adelaide, Adelaide, SA, Australia

The glycolytic enzyme pyruvate kinase M2 (PKM2) is a target gene of the transcription factor hypoxia inducible factor-1 (HIF-1), and in turn acts as a coactivator for HIF-1 in cancer cells. PKM2 enhances DNA binding and coactivator recruitment to hypoxia response elements of HIF-1 target genes, increasing transcription. This leads to various cellular responses including upregulation of glycolysis. Together PKM2 and HIF-1 have been implicated in driving aerobic glycolysis (also known as the Warburg effect) in cancer and other proliferating cells. Surprisingly the non-proliferating mammalian retina also displays the Warburg effect. We hypothesise that PKM2 and HIF-1 influence metabolic gene transcription and drive the Warburg effect in the retina, similar to cancer. Here we show PKM2 and HIF-1 expression in Müller glial cells of the retina, and demonstrate that the retina *in vivo* expresses PKM2 [1]. We investigate the role of PKM2 in cellular metabolism and metabolic gene expression in Müller cells, using *in vitro* knockdown and CRISPR/Cas9-mediated knockout models. To facilitate studies of Müller cells, we have generated a novel spontaneously immortalised rat Müller cell line SIRMu-1. We profile the transcriptome of these cells by RNA sequencing and compare this to primary Müller cells and the SV40-immortalised Müller cell line rMC-1. We also report bioinformatic analysis of the differential expression of metabolic genes among retinal cell types from published single-cell RNA sequencing data. Understanding the molecular basis of transcriptional control of retinal metabolism may aid diagnosis and

treatment of retinal diseases. Additionally, if mechanisms underlying the Warburg effect in cancer and the retina are conserved, caution needs to be exercised when designing novel cancer treatments targeting aerobic glycolysis as they may also damage the retina.

1. [1] Casson R.J., Wood J.P., Han G., Kittipassorn T., Peet D.J., Chidlow G. (2016) Invest. Ophthalmol. Vis. Sci. 57(1):66-80.

173

Regulation of novel 'eosinokines' in the development of beige fat

Alexander Knights[1], Emily Vohralik[1], Peter Houweling[2], Elizabeth Stout[1], Kyle Hoehn[1], Merlin Crossley[1], Kate Quinlan[1]

1. UNSW Sydney, Kensington, NSW, Australia

2. Murdoch Children's Research Institute, Parkville, VIC, Australia

Fat, or adipose tissue, comes in two main forms – white, where energy is stored, and brown, where energy is burned in a process called thermogenesis to produce heat. In the last decade it has been discovered that white fat, where energy is stored in the form of lipid droplets, can be naturally converted to become like brown fat, where energy is burned. This phenomenon is called 'browning' and results in the development of 'beige' fat. This is a key area of current research into treating metabolic diseases because the induction of beige fat results in weight loss and improved metabolic parameters. Our research is focused on how this transition is regulated at the transcriptional level, with a particular interest in the contribution of resident immune cells.

While studying a mouse model lacking the transcriptional repressor Krüppel-like factor 3 (KLF3), we noticed that these mice remain lean and glucose tolerant on a Western diet, with greatly reduced fat mass. Upon closer examination, we determined that the white fat of KLF3 knockout mice appeared 'browner' due to increased mitochondria, had more resident eosinophils and showed high levels of thermogenic gene expression – all distinct hallmarks of beige fat. We isolated eosinophils, where KLF3 is highly expressed, from the fat of wildtype and KLF3 knockout mice and found that cells lacking KLF3 expressed important secreted factors implicated in beige fat development. These results point towards a role for KLF3 in the regulation of important eosinophil-derived factors that drive the 'browning' of fat, which we have termed 'eosinokines'.

Initial studies have shown that the administration of certain eosinokines can improve whole body energy expenditure and metabolism, alleviating metabolic disorders. Further understanding of how these eosinokines are regulated will help us to better tailor for them therapeutic delivery.

174

Investigating the role of mutant DNMT3A in Acute Myeloid Leukaemia and other cancers

Erin Lawrence[1, 2], Martin Pal[3, 2], Andrew Kueh[3, 2], Lin Tai[3, 2], Andreas Strasser[3, 2], Andrew Wei[4], Marco J Herold[3, 2]

1. Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia

2. Department of Medical Biology, The University of Melbourne, Melbourne, Vic, Australia

3. Walter and Eliza Hall Institute of Medical Research, Melbourne, Vic, Australia

4. Australian Centre for Blood Diseases, Melbourne, Vic, Australia

Acute Myeloid Leukaemia (AML) is characterized by aberrant proliferation of clonal myeloid stem and progenitor cells. Mutations in the DNA Methyltransferase 3a (DNMT3A) and other epigenetic regulators have been identified as initiating events in AML. DNMT3A is mutated in ~20% of AMLs making it one of the most frequently mutated genes. Therefore, we have used CRISPR/Cas9 technology to introduce the most prevalent mutation found in human DNMT3a in mice by mutating R878 (murine homologue of R882) to histidine. The $DNMT3A^{mut}$ mice do not spontaneously develop AML, which is consistent with previous studies that suggest that other oncogenic drivers, such as $NRAS^{mut}$ or $IDH2^{mut}$, are required for malignant transformation. Therefore, we have transduced haematopoietic stem and progenitor cells (HSPCs) derived from $DNMT3A^{mut}$ and wildtype foetal liver cells with retroviral constructs encoding for $NRAS^{G12D}$ and $IDH2^{R140Q}$ and transplanted them into lethally irradiated recipient mice. While co-expression of $IDH2^{mut}$ in $DNMT3a^{mut}$ foetal liver cells led with very low penetrance and long latency (~300 days) to increased HSC frequencies and first signs of malignancy, $NRAS^{mut}$ co-expression induced T-ALL in a high proportion of mice transplanted with $DNMT3a^{mut}$/ $NRAS^{mut}$ foetal liver cells. We are currently establishing cell lines from the different $DNMT3a^{mut}$ cancer cells allowing us to compare the epigenetic landscape between wildtype and $DNMT3a^{mut}$ cells. In order to test, whether $DNMT3a^{mut}$ is required for the sustained growth of the malignant cells, we have introduced an additional sgRNA binding site into the mutant DNMT3A allele. By targeting this allele, we can effectively delete the mutant DNMT3A in the cancer cells and track the survival of leukaemic

cells *in vitro* and *in vivo.* These experiments will be combined with standard chemotherapy and targeted therapies to find potential treatment regimens useful in the clinic for patients presenting with DNMT3a^{mut} cancers.

175

Mitochondrial variant calling in 15,496 individuals in the Genome Aggregation Database (gnomAD)

Kristen Laricchia, Grace Tiao, gnomAD Consortium, Daniel MacArthur, Vamsi Mootha, Sarah Calvo, Monkol Lek

The Genome Aggregation Database (gnomAD; http://gnomad.broadinstitute.org) contains 15,496 whole genome sequences of unrelated individuals that are used to aid variant interpretation for human diseases. However, variants in these genomes have currently only been called for the 22 autosomes and chromosome X, omitting variants in the mitochondrial genome (mtDNA). Mitochondria are maternally inherited organelles each containing many copies of the 16,569bp mtDNA. mtDNA encodes 13 proteins all involved in oxidative phosphorylation, which is required for generating ~90% of all cellular ATP. Over 400 mutations in mtDNA have been associated with human disease, and identifying all mtDNA variants in thousands of gnomAD genomes would provide an excellent resource to help researchers assess the frequency and interpretation of mitochondrial variants.

With the goal of including mtDNA variants in future releases of gnomAD, we have started a pilot study to identify and assess mtDNA variants in over 15,000 gnomAD genomes. The mtDNA is present in thousands of copies per cell, and the gnomAD genomes typically have >5000x mtDNA coverage. However, a specialized variant calling methodology is required because mtDNA variants can exist at any level of heteroplasmy (percent of mtDNA molecules with variant). To call mitochondrial variants, we ran mtDNA-Server (Weissensteiner et al. 2016), a tool that is optimized for identifying mitochondrial variants, detecting contamination, and determining haplogroups (maternal-line ancestry). We identified low heteroplasmic, high heteroplasmic, and homoplasmic variants among gnomAD individuals. To perform quality control, we assessed maternal transmission rates in mother-child pairs and in sample duplicates. All individuals were also classified into haplogroups, which aids in detecting contamination and provides additional information on ancestry to complement population assignment using the nuclear genomes. Including mtDNA variants in future releases of the gnomAD database will provide critical population frequency information to help researchers understand the role of mitochondrial variation in health and disease states.

176

Personalized analytics to improve diagnostic rates in clinical sequencing.

Timo Lassmann[1], Denise Anderson [1]

1. Telethon Kids Institute, Subiaco, WESTERN AUSTRALIA, Australia

'Personalized' or 'precision' medicine has been heralded as the next big revolution in healthcare. With the cost of DNA sequencing now low enough for whole genome sequencing (WGS) to be considered a standard diagnostic tool, WGS is expected to replace targeted gene and/or exome sequencing as the default protocol within the next five years. The major bottleneck remains clinical interpretation of the data - identifying the disease-causing variant or variants among thousands of benign or functionally unrelated variants.

Here I will introduce a new approach integrating a patient's phenotype with public 'whole body' gene expression data and sets of known disease annotations to generate patient specific prediction models.

177

Elucidating the molecular signatures underlying ALT activity in tumours

Michael Lee[1], Erdahl T Teber[2], Oliver Holmes[3], Katia Nones[4], Ann-Marie Patch[4], Rebecca A Dagg[5], Loretta MS Lau[5], Joyce H Lee[5], Christine E Napier[5], Jonathan W Arthur[2], Sean M Grimmond[6], Nicholas K Hayward[7], Peter A Johansson[7], Graham J Mann[8], Richard A Scolyer[9], James S Wilmott[9], Roger R Reddel[5], John V Pearson[4], Nicola Waddell[3], Hilda A Pickett[10]

1. Children's Medical Research Institute, Westmead, NSW, Australia

2. Bioinformatics Unit, Children's Medical Research Institute, University of Sydney, Westmead, NSW, Australia

3. Genome Informatics Group, QIMR Berghofer Medical Research Institute, Herston, Queensland, Australia

4. Medical Genomics Group, QIMR Berghofer Medical Research Institute, Herston, Queensland, Australia

5. Cancer Research Unit, Children's Medical Research Institute, University of Sydney, Westmead, NSW, Australia

6. Centre for Cancer Research, University of Melbourne, Melbourne, Victoria, Australia

7. Oncogenomics Group, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia

8. Centre for Cancer Research, Westmead Institute for Medical Research, Westmead, NSW, Australia

9. Discipline of Pathology, Sydney Medical School, The University of Sydney, Sydney, NSW, Australia

10. Telomere Length Regulation Unit, Children's Medical Research Institute, University of Sydney, Westmead, NSW, Australia

The replicative immortality of human cancer cells is achieved by activation of a telomere maintenance mechanism (TMM). To achieve this, cancer cells utilise either the enzyme telomerase, or the Alternative Lengthening of Telomeres (ALT) pathway. These distinct molecular pathways are incompletely understood with respect to activation and propagation, as well as their associations with clinical outcomes. We have developed a pipeline capable of accurately quantitating telomere variant repeats from whole genome sequencing (WGS) data. We validated TMM in two tumour sample sets comprising of 86 pancreatic neuroendocrine tumours and 81 melanomas. We then used our telomere variant repeat pipeline to generate a random forest classifier, which was capable of predicting TMM with 91.6% accuracy using telomere length and sequence content alone. This classifier was then applied to 908 WGS datasets across 22 tumour types, for which information regarding the activated TMM was not previously known. We found that aberrations in TMM associated genes, such as loss-of-function mutations in ATRX/DAXX and TERT promoter mutations, were not common across all tumour types nor mutually exclusive to one TMM. Analysis of pathways that were correlated with the activation of ALT identified autophagy and cell cycle control of chromosomal replication pathways as important for the survival of ALT tumours. Our analysis also revealed extensive genetic heterogeneity between tumour types, highlighting the complexities involved in wide scale genome studies of TMMs. Our approach has the potential to be further developed with the addition of more TMM validated datasets and can be applied to larger tumour datasets in order to elucidate the tumour type specific mechanisms of ALT activation.

178

Stepwise engagement of the Killer T cell function is underscored by H3K27 demethylation

Jasmine Li[1], Moshe Olshansky[1], Brendan Russ[1], Kristine Hardy[2], Sudha Rao[2], Stephen Turner[1]

1. Monash University, Clayton, Victoria, Australia

2. Department of Molecular and Cellular Biology, University of Canberra, Canberra, ACT, Australia

The acquisition of cytotoxic functions by effector CD8+ T cells (CTL) is a critical component of the adaptive immune system as they provide targeted antigenic responses to infections. This differentiation process, which requires the integration of T cell and co-stimulatory signaling cascades, is underscored by epigenetic re-programming, where coordinated changes to chromatin structure and histone modifications allow establishment of CTL transcriptional signatures. Recently, the regulation of a specific repressive histone modification (H3K27me3) has been shown to influence naïve to effector/memory differentiation in CD8+ T cells. Here, we track the dynamics of H3K27me3 demethylation during the early phase of T cell activation against Influenza A in an *in vitro* mouse model. Using ChIP- and RNA-sequencing, we demonstrate that early-activated T cells engage in a series of stepwise transcriptional changes, with transcription factors critical for T cell differentiation such as *Tbx21* (encoding for T-bet) and *Irf4* being expressed at as early as 3 hours post-activation. Importantly, the use of a demethylase inhibitor demonstrated that the magnitude of their transcription responses was dependent on H3K27me3 demethylation, indicating that the removal of this repressive histone modification significantly shapes early cytotoxic gene expression programs through the activation of these transcription factors. Understanding these molecular signatures therefore represents a novel regulatory step involved in the formation of effector and memory CD8+ T cells.

179

A Cross Species and Multi-omic Analysis Indicates the Exon/Intron Structure Preceding Initiation Codons Predicts their Translation Efficiency

Chun Shen Lim[1], Samuel Taylor-Wardell[1], Torsten Kleffmann[1], Christopher Brown[1]

1. University of Otago, Dunedin, OTAGO, New Zealand

Introns are distributed unevenly along genes — common in the coding sequence but rare in the 3′ untranslated region (3′UTR). It is proposed that the positions of introns are maintained by natural selection because introns in the 3′UTRs can reduce expression through nonsense-mediated mRNA decay[1]. In contrast, introns in the mRNA leaders of complex eukaryotes are common but often overlooked[2]. These introns are spliced out before

translation, leaving exon-exon junctions in the mRNA leaders (5′EEJ). A cross species and multi-omic approach was taken to study exon structure of the mRNA leaders, inferring functions and selective constraints. Among the features of mRNA leaders examined, the number of 5′EEJ is inversely correlated with the main protein translation, as does the number of uORFs. Across the five species studied, the lowest levels of translation were observed for mRNAs with both 5′EEJ and uORFs. This class of mRNAs also have ribosome footprints on uORFs, with strong triplet periodicity indicating uORF translation. Comparative genomic analysis shows that the positions of both 5′EEJ and uORFs are conserved in human-mouse divergence. Thus the uORFs of the mRNAs with 5′EEJ are more likely to be functional and have regulatory roles.

1. Hellens, R. P., Brown, C. M., Chisnall, M. A. W., Waterhouse, P. M., & Macknight, R. C. (2016). The emerging world of small ORFs. Trends in Plant Science, 21(4), 317-328. doi: 10.1016/j.tplants.2015.11.005
2. Chung, B. Y. W., Simons, C., Firth, A. E., Brown, C. M., & Hellens, R. P. (2006). Effect of 5'UTR introns on gene expression in Arabidopsis thaliana. BMC Genomics, 7, 120. doi: 10.1186/1471-2164-7-120

180

Nuclear factor one transcription factors regulate developmental enhancers during brain development: insights from motif discovery

Jonathan WC Lim[1], Jens Bunt[1], Linda J Richards[1, 2]

1. Queensland Brain Institute, The University of Queensland, St Lucia, QLD, Australia

2. School of Biomedical Sciences, The University of Queensland, St Lucia, QLD, Australia

The generation of specific cell types in a timely and orderly manner is a prerequisite for the formation of neural circuits during development. This process is largely governed by sequence-specific transcription factors, which bind to their cognate motifs on DNA to modulate gene expression. One such family of transcription factors known to regulate important developmental processes is the Nuclear factor one (NFI) family of transcription factors. Our analyses of *cis*-regulatory elements that were previously identified by the Mouse ENCODE project suggest that these transcription factors play a critical role in neuronal differentiation during cortical development. Specifically, we observed that the dimeric NFI motif, characterised by the palindromic sequence TTGGC$(N)_5$GCCAA, is highly enriched at developmental enhancers derived from the embryonic forebrain. Our analyses demonstrate that the NFI motif is also the most abundant and most highly enriched transcription factor motif at enhancers during early neurogenesis – between E11.5 and E13.5 in the developing cortex. Ongoing analyses of the epigenome in *Nfi* knockout mouse models have demonstrated that these transcription factors contribute towards enhancer activation during cortical development. Gene ontology reveals a strong correlation between epigenomic changes and the developmental phenotypes observed in knockout mice and humans with *NFIB* haploinsufficiency. In summary, our work identifies a major regulator of enhancer activity during cortical development, and contributes towards identifying the downstream regulatory networks critical for neurogenesis in the developing brain.

181

Investigating computational analysis pipelines and genomic proximity interactions in T lymphocytes

Ning Liu[1, 2], Tim Sadlon[3, 1], Stephen Pederson[2], Simon Barry[3, 1], James Breen[1, 2]

1. Robinson Reseach Institute, University of Adelaide, Adelaide, SA, Australia

2. Bioinformatics Hub, University of Adelaide, Adelaide, SA, Australia

3. Division of Paediatric Medicine, Women's and Children's Hospital, Adelaide, SA, Australia

Chromosome Conformation Capture (3C) technology is a method used for investigating three-dimensional (3D) genome structure, whereby segments of a genome that are in close-proximity can be identified and used to infer their spatial relationship. A 3C-derived method, High-resolution Chromosome Conformation Capture sequencing (HiC-seq) have been used to identify genes that can be affected by distal interactions such as long-range promoter-enhancer contacts that interact with immune system regulators. Although HiC-seq has been widely used to identify 3D interactions genome-wide in many species, many of the analysis tools have yet to be critically assessed. Here, we used publically available HiC-seq data to investigate and compare three major steps of HiC-seq data analysis workflow, including raw HiC-seq data processing, topologically-associated domains (TADs) identification algorithms and visualisation tools. We then applied our validated toolset to a DNaseI-treated, HiC-seq dataset sampled from human conventional T cells (Tconv cells) to investigate the ability of the tools at analysing relative low-coverage datasets. Whilst HiC-seq data analysis requires a significant sequencing coverage, applying HiC-Pro, an insulation score algorithm for TAD identification and HiCPlotter for visualisation, we identified a total of 4,818,855 long-range interactions, leading to the prediction of 3275 TADs genome-wide. Using this HiC-seq data along with other conformation assays

(i.e. 4C-seq), we show that an upstream super-enhancer and promoter of the master T cell regulator *SATB1* are located within the same TAD region, supporting the hypothesis that long-range interactions regulate the function of *SATB1*, and that sequence variants in enhancer elements may effect the pathogenicity of autoimmune diseases.

182

Quantifying chromatin structure at the level of nucleosomes by correlative live cell FLIM-FRET microscopy and fixed cell single molecule localization microscopy.

Jieqiong Lou[1], Elizabeth Hinde[1]

1. University of Melbourne, Melbourne, VICTORIA, Australia

A key emerging contributor to genome function is the architectural organisation of the cell nucleus. DNA is wrapped around histones to form nucleosomes, which are folded into chromatin fibres and looped into a structure that eventually becomes a three-dimensional chromosome. The rearrangements in chromatin structure that control DNA template access occur on a spatiotemporal scale that render them 'invisible' to live cell imaging. Thus, in recent work we probed chromatin compaction at the level of nucleosomes by fluorescence lifetime based microscopy of FRET between fluorescently labelled histones. Here we multiplex this methodology with single localization microscopy of histone post-translational modifications, to determine the molecular mechanism that regulates chromatin dynamics at a DNA damage site with unprecedented spatial resolution.

201

ELEVATED FOETAL GLOBIN CAUSED BY NATURAL REGULATORY MUTATIONS THAT DISRUPT FOETAL REPRESSOR BINDING

Gabriella E Martyn[1], Beeke Wienert[1], Lu Yang[1], Manan Shah[1], Alister PW Funnell[1], Kate GR Quinlan[1], Merlin Crossley[1]

1. School of Biotechnology and Biomolecular Sciences, UNSW Australia, Sydney, NSW, Australia

Beta-haemoglobinopathies, such as sickle cell disease and beta-thalassaemia are caused by mutations in the adult *beta-globin* gene. They are amongst the most commonly inherited disorders, with the World Health Organisation estimating that at least 5% of adults are carriers for a haemoglobin disorder. With current therapeutic strategies carrying various limitations, reactivating the developmentally silenced foetal *globin* gene is an attractive approach for treating beta-haemoglobinopathies.

A rare group of individuals with a benign condition called the Hereditary Persistence of Foetal Haemoglobin (HPFH), continue to express foetal haemoglobin into adult life. Co-inheriting a HPFH mutation with a beta-haemoglobinopathy mutation leads to alleviated symptoms of the disease. HPFH is caused by single point mutations within the promoter of the foetal *globin* gene. The -115 site of the foetal *globin* promoter contains four reported HPFH mutations and a small 13 base pair deletion. These point mutations are hypothesised to disrupt the binding of a transcriptional repressor which would normally silence the foetal *globin* gene around the time of birth. Our aim was to identify this repressor.

We studied a range of potential transcription factors to identify a DNA-binding domain which could directly bind to the -115 site of the foetal *globin* promoter *in vitro* using electrophoretic mobility shift assays (EMSA). We compared binding with and without the HPFH mutations and identified BCL11A as a candidate repressor. Furthermore, introduction of the naturally occurring HPFH mutations into erythroid cells by CRISPR/Cas9 genome editing disrupts BCL11A binding and raise foetal *globin* expression. These results identify BCL11A as the repressor which binds to the -115 site of the foetal *globin* promoter and reveal the mechanism of how the HPFH mutations operate, providing insights into new potential therapeutic targets for treating beta-haemoglobinopathies.

202

Cross-regulation between the ethylene and abscisic acid hormone signalling pathways is regulated by EDF transcription factors

Mathew G. Lewsey[1], Joseph R Ecker[2]

1. La Trobe University, Bundoora, VIC, Australia

2. Salk Institute for Biological Studies-HHMI, La Jolla, CA, USA

Several plant hormone signalling pathways exist to co-ordinate the diverse aspects of growth, development and defence. They are responsive to the changing internal and external conditions experienced during the

lifespan of plants. Extensive cross-regulation occurs between these hormone signalling pathways, so as to generate appropriate responses. This cross-regulation frequently occurs by synergistic or antagonistic regulation of shared sets of genes, through the activity of hormone-responsive transcription factors. We have studied the transcription factors that regulate responses to the hormone ethylene. Ethylene is involved in diverse processes including fruit ripening, growth in the dark and defence. EIN3 is the primary transcription factor that regulates gene expression in response to ethylene, targeting and up-regulating hundreds of genes. We demonstrate that EIN3 induces expression of three repressive transcription factors from the EDF family. These in turn target hundreds of genes for down-regulation during the response to an ethylene stimulus. The targets of the activatory EIN3 are predominantly known ethylene response components. However, the targets of the repressive EDFs are enriched for components of other hormone signalling pathways, particularly the abscisic acid pathway. Ethylene-abscisic acid has long been known to exist, but the molecular mechanisms have not been characterised previously. Our results illustrate that the EDFs may be the mechanism by which ethylene suppresses abscisic acid signalling, and that this results in measurable effects on plant photosynthetic capacity. Moreover, we show that plant hormone signalling involves a network of responsive transcription factors acting in a co-ordinated fashion to condition the ultimate gene expression responses.

203

RNA-dependent epigenetic silencing directs transcriptional down regulation caused by intronic repeat expansions

Hannes Eimer[1], Sridevi Sureshkumar[1], Champa Bandaranayake[1], Andrei Seleznev[1], Calvin A Kraupner-Taylor[1], Avilash Singh Yadav[1], Tamblyn Thomason[1], Stephen Fletcher[2], Stephanie Francis Gordon[1], Bernard J Carroll[2], Sureshkumar Balasubramanian[1]

1. Monash University, Clayton, VIC, Australia

2. The University of Queensland, Brisbane

Transcriptional down regulation caused by intronic triplet repeat expansions underlies diseases such as Friedreich's ataxia. This down regulation of gene expression is coupled with epigenetic changes but the underlying mechanisms are unknown. Here, we show that the TTC/GAA triplet expansion at the IIL1 gene of Arabidopsis thaliana results in accumulation of 24-nt siRNAs and repressive histone marks at the IIL1 locus, which in turn causes its transcriptional down regulation and an associated phenotype. Knocking down DICER LIKE-3 (DCL3), which processes 24-nt siRNAs, suppressed triplet expansion associated phenotype and helped overcome this transcriptional down regulation and epigenetic silencing. Furthermore, knocking down HEN1, AGO4 or NRPE1, additional components of the siRNA-dependent epigenetic silencing pathway, also suppressed both the repeat expansion phenotype and transcriptional down regulation of IIL1. Thus our results show that triplet repeat expansions can lead to local siRNA biogenesis, which in turn down regulates transcription through an AGO4-dependent epigenetic modification.

204

Duplication and deletion of key *SOX9* enhancers causes sex reversal in humans

Brittany Croft, Thomas Ohnesorg, Josephine Bowels, Emanuele Pelosi, Vincent Harley, Peter Koopman, Andrew Sinclair

Disorders of Sex Development (DSDs) encompass a wide spectrum of conditions and often manifest with atypical gonads or genitalia. The majority of DSD patients cannot be given an accurate diagnosis, which severely comprises their clinical management. While mutations in coding regions of gonad genes have been important in understanding the aetiology of DSD little attention has focussed on the regulatory regions of gonad genes.

It has been 27 years since the Y-linked *SRY* gene and its role as the master of mammalian sex determination were discovered. Studies in mice implicate Sox9 as a direct target of Sry, and while *SOX9* upregulation in the early gonad is crucial for testis development, the *SOX9* sequences bound by SRY during human testis determination have not been identified. We analysed copy number variations (CNVs) in the upstream regulatory region of *SOX9* in DNA from patients with DSD, allowing us to define several minimal critical regions for sex-reversal.

Using a combination of luciferase tiling and bioinformatics approaches we identified three novel human testis enhancers 5' of *SOX9*. In cell-based reporter assays these enhancers responded to different combinations of testis specific regulators including SRY, NR5A1 (SF1) and SOX9 itself. Enhancers that showed the strongest activity *in vitro* were used to generate transgenic mice. *In vivo*, two enhancers drove reporter gene expression in mouse embryonic gonads, while mice null for the third enhancer showed reduced *Sox9* transcription. Given that duplications and deletions of these enhancers in DSD patients result in sex reversal (*SRY* negative, 46,XX males or 46,XY females respectively) our results suggest a mechanism by which these upstream enhancers

play crucial roles in human sex development. Together these three enhancers may provide the missing link between *SRY* and *SOX9* initiation, upregulation and maintenance – all fundamental to testis development.

205

Canonical PRC2 function is essential for mammary gland development and affects chromatin compaction in mammary organoids

Ewa M Michalak[2, 1], Michael J.G Milevskiy[1], Rachel M Joyce[1], Johanna F Dekkers[1], Paul R Jamieson[1], Bhupinder Pal[1], Yifang Hu[1], Stuart H Orkin[3], Warren S Alexander[1], Geoffrey J Lindeman[1], Gordon K Smyth[1], Jane E Visvader[1]

1. Walter and Eliza Hall Institute, Melbourne, Victoria, Australia

2. PETER MACCALLUM CANCER CENTRE, Melbourne, VIC, Australia

3. Dana Farber Cancer Instiute, Boston, USA

Distinct transcriptional states are maintained through organization of chromatin, resulting from the sum of numerous repressive and active histone modifications, into tightly packaged heterochromatin versus more accessible euchromatin. PRC2 is the main mammalian complex responsible for Histone 3 lysine 27 tri-methylation (H3K27me3), and is integral to chromatin organization. Using *in vitro* and *in vivo* studies, we show that deletion of *Suz12*, a core component of all PRC2 complexes, results in loss of H3K27me2/me3, completely blocks normal mammary gland development and profoundly curtails progenitor activity in 3D organoid cultures. Through the application of mammary organoids to bypass the severe phenotype associated with *Suz12* loss *in vivo*, we have explored gene expression and chromatin structure in wildtype and *Suz12*-deleted basal-derived organoids. Analysis of organoids led to the identification of lineage-specific changes in gene expression and chromatin structure, inferring cell-type specific PRC2-mediated gene silencing of the chromatin state. These data indicate that canonical PRC2 function is essential for development of the mammary gland through the repression of alternate transcription programs and maintenance of chromatin states.

206

Welcome to the macrogenomics era: a long read revolution

Martin Smith[1], Kirston Barton[1], James Ferguson[1], Ruth Lyons[1], Shaun Carswell[1], David Thomas[1], Vanessa Hayes[1], Marcel Dinger[1]

1. Garvan Institute of medical research, Darlinghurst, NSW, Australia

Short read shotgun sequencing has infiltrated nearly all genomics research programs thanks its expansive depth of coverage and great accuracy. However, its technical shortcomings leave several aspects of genomics unresolved, such as the extend and impact of complex structural variations and repetitive regions on genome biology.

The advent of long read sequencing promises to resolve these shortcomings. Once stigmatised for its high error-rate, Oxford nanopore technologies can produce ultra long reads while detecting base modifications, in real time. As read lengths greater than 100kb are becoming increasingly routine, we can observe entire chunks of native human chromosomes at single nucleotide resolution. Here, we illustrate how ultra-long reads can shed light into macrogenomics by exposing the sequence components of chromothrypsis-derived neochromosomes, hyper-methylated regions, and the first megabase long sequencing read.

With regards to nanopore's base calling error-rate, we relate how signal-level algorithms can lead to highly accurate sequence analysis, such as single cell barcode demultiplexing. By venturing into ‘squiggle space’, we describe how electronic signal from nanopore sequencers can be harnessed to increase the amount of de-multiplexed full length cDNA sequences from 10X Chromium single-cell sequencing by more than 2-fold. The unique nature of nanopore sequencing, and the bioinformatics opportunities it presents, mark an exciting new era for genomic research.

207

Molecular and cellular role of RNA-binding proteins in cardiac biology and disease

Lithin Louis[1], Nikolay Shirokikh[1], Yalin Liao[2], Meghna Sobti[3], Nicola J Smith[3], Thomas Preiss[1]

1. John Curtin School of Medical Research, Acton, ACT, Australia

2. School of Veterinary and Life Sciences, Murdoch University, Murdoch , WA, Australia

3. Victor Chang Cardiac Research Institute, Darlinghurst, NSW, Australia

RNA-Binding Proteins (RBPs) are omnipresent partners of cellular RNA. For example, eukaryotic mRNAs dynamically associate with RBPs assisting their transcription, maturation, nuclear to cytoplasmic translocation,

localization, translation and degradation. The 'RNA interactome capture' method uses *in-vivo* UV crosslinking followed by oligo(dT) purification and mass spectrometry of enriched peptides. It has led to the identification of hundreds of novel RBPs in cardiomyocytes (Liao *et al.*, *Cell Rep* 2016, **16:**1456–1469). This included multiple metabolic enzymes, given further credence to the RNA-Enzyme-Metabolite (REM) hypothesis, which proposes the existence of regulatory links between gene expression and intermediary metabolism mediated by RNA-binding metabolic enzymes. Possible functions include enzymes 'moonlighting' in the regulation of mRNA utilisation or (noncoding) RNAs regulating enzyme function (Castello *et al.*, *Trends Endocrinol Metab* 2015, **26:**746–757).

To explore these scenarios, we focus on prominent cardiomyocyte metabolic RBPs such as enolase, citrate synthase, aconitase-2, malate dehydrogenase and isocitrate dehydrogenase. We are in the middle of implementing a plan that involves the use of *in vivo* UV or formaldehyde crosslinking to stabilise the REM partners, followed by affinity purification *via* a GFP tag on the enzyme and identification of RNA binders by deep sequencing; precise binding sites in the RNA will be located by limited degradation and/or reverse transcription stalling approaches. We will further use site-specific mutagenesis of both target RNA and enzyme in HL-1 cells and mice hearts at different developmental stages. Given the significance of metabolic pathways in heart, unravelling REM interaction networks can provide new insights into cardiomyocyte development, biology, and pathophysiology.

208

FANCM protects telomeres from aberrant replication fork processing in cancers engaging in Alternative Lengthening of Telomeres (ALT)

Robert Lu[1], Alexander Sobinoff[1], Joshua allen[1], Christopher Tomlinson[1], Michael Lee[1], Hilda Pickett[1]

1. Children's Medical Research Institute, Westmead, NSW, Australia

Telomeres are terminal TTAGGG repeats that cap chromosome ends and their shortening is the basis of replicative senescence. Cancer cells must activate a telomere maintenance mechanism (TMM) to overcome this barrier. 10-15% of cancers utilise a telomerase-independent mechanism termed Alternative Lengthening of Telomeres (ALT). Patients with ALT tumours often have a poorer prognosis compared to telomerase-positive counterparts and in need of targeted treatments. A potential target is the susceptibility of telomeres to replication stress, which ALT-cancers exploit to undergo recombinational telomere synthesis and repair. There has been mounting evidence implicating a major role for Fanconi Anemia (FA) pathway proteins at telomeres. The FA pathway is canonically required for inter-strand cross-link repair and is involved in the protection of stalled replication forks. We have identified a role for the key sensing member of the FA pathway, FANCM, in maintaining telomere integrity and regulating ALT activity. Specifically, we found that loss of FANCM induced telomere dysfunction in both ALT and telomerase cells. In ALT cells, we observed a "hyper-ALT" phenotype, characterised by a dramatic increase in telomere length heterogeneity, ALT-associated PML bodies and extrachromosomal telomeric DNA species including C-circles and G-circles. This is driven by endonuclease-mediated cleavage of stalled replication forks at telomeres and coincides with upregulation of the pRPA-S4/S8-ATR and DNA-PK DNA-damage pathways. We have further determined that the MM2 and translocase domains of FANCM are functionally required for telomere protection. Overall, we have identified a potential vulnerability of ALT cancers, through FANCM deficiency, which may be exploited in the future by synthetic lethal targeting of proteins involved in DNA-damage repair.

209

Investigating the role of ZBTB7A homodimerisation in adult erythroid cells

Lana C Ly[1], Kate GR Quinlan[1], Merlin Crossley[1]

1. School of Biotechnology & Biomolecular Sciences, UNSW Sydney, Kensington, NSW, Australia

β-haemoglobinopathies such as sickle-cell anaemia and β-thalassemia are among the world's most common monogenic diseases. These diseases are associated with mutations within the *β-globin* genes that code for adult haemoglobin, resulting in impaired oxygen transport. Current treatments to alleviate symptoms are limited and only offer short-term relief with lifelong administration required. As 5% of the world population are carriers for haemoglobinopathies, a gene therapeutic approach to cure these diseases may be an effective approach. It has been shown that in individuals with a β-haemoglobinopathy, elevation of foetal haemoglobin levels leads to ameliorated symptoms. Thus, reactivation of the developmentally silenced foetal *γ-globin* genes offers a promising approach to treat β-haemoglobinopathies.

Key transcriptional regulators have been identified in recent years that control the developmental switch between foetal and adult haemoglobin. The aim of our research is to explore a transcriptional repressor called Zinc Finger and BTB Domain Containing 7A (ZBTB7A) that has been discovered to repress the foetal *γ-globin* genes. Knocking out the *Zbtb7a* gene in an erythroid adult-like cell model dramatically increases *γ-*

globin levels. While the mechanism of repression is still unknown, we hypothesise that ZBTB7A's ability to homodimerize via its BTB/POZ domain plays a key role in its binding to the foetal *γ-globin* genes to mediate *γ-globin* repression in adulthood.

Therefore, this study aims to investigate the functional importance of ZBTB7A homodimerisation. One amino acid within the BTB/POZ domain of ZBTB7A has been shown to be crucial for homodimerisation. By introducing an amino acid substitution into endogenous *Zbtb7a* using CRISPR/Cas9-mediated genome editing in an erythroid cell model, we aim to abrogate ZBTB7A homodimerization and elicit the mechanism and role in its ability to transcriptionally repress genes, with a particular focus on *γ-globin* repression.

210

KLF1 is a pioneer transcription factor for erythroid cells

Graham W Magor[1], Kevin R Gillinder[1], Charles C Bell[2], Darren Korbie[3], Matt Trau[3], Andrew C Perkins[1, 4]

1. Monash University, Melbourne, Victoria, Australia

2. Peter MacCallum Cancer Research Institute, Melbourne, Victoria, Australia

3. Australian Institute of Biotechnology and Nanotechnology, Brisbane, Queensland, Australia

4. The Alfred Hospital, Melbourne, Victoria, Australia

Krüppel-like-factor 1 (KLF1) is a transcription factor (TF) expressed uniquely within the red blood cell lineage. KLF1 binds to the promoters and/or enhancers of many genes critical to erythropoiesis and activates their expression. Pioneer factors are a class of TFs that are able to interact with 'closed' DNA that is wound around nucleosomes. They 'pioneer' the way for other factors to bind by re-positioning nucleosomes, exposing the hitherto occluded DNA. The relative permissiveness of DNA loci can be measured using ATAC-seq (Assay for Transposase-Accessible Chromatin) which relies on the DNA being accessible (i.e. nucleosome free) to a Tn5 transposase loaded with next-generation sequencing adaptors.

Using a tamoxifen inducible version of mouse KLF1 (Gillinder et al., 2017), we demonstrate that the KLF1 protein possesses pioneering activity. Upon induction we find KLF1 occupies ~3800 genomic regions. The majority show an increase in chromatin accessibility over the parental KLF1 null cell line, although many are already 'open' prior to KLF1 induction. Many of the induced sites are at well-known erythroid enhancers such as the -26kb enhancer of the alpha-globin LCR (Hay et al., 2016). The zinc finger DNA-binding domain (DBD) of KLF1 binds DNA well *in vitro* and *in vivo* but is unable to alter chromatin *in vivo*, suggesting the latter function is encoded in the non-DBD. The mechanism by which KLF1 opens chromatin is an active area of investigation and likely relies upon interaction with components of the chromatin-remodelling SWI/SNF complex.

This work establishes KLF1 as a *bone fide* pioneer TF and begins to dissect the mechanism of its action. As KLF1 is the founding member of the KLF/SP superfamily, this work is likely to establish a paradigm for the function of this TF family in gene regulation and reprograming.

1. Gillinder, Kevin R; Ilsley, Melissa D; Nébor, Danitza; Sachidanandam, Ravi; Lajoie, Mathieu; Magor, Graham W; Tallack, Michael R; Bailey, Timothy; Landsberg, Michael J; Mackay, Joel P; Promiscuous DNA-binding of a mutant zinc finger protein corrupts the transcriptome and diminishes cell viability Nucleic acids research 45 3 1130-1143 2017
2. Hay, D., Hughes, J.R., Babbs, C., Davies, J.O., Graham, B.J., Hanssen, L.L., Kassouf, M.T., Oudelaar, A.M., Sharpe, J.A., Suciu, M.C. and Telenius, J., 2016. Genetic dissection of the [alpha]-globin super-enhancer in vivo. Nature genetics.

211

Restoring Central Inhibition Relieves Chronic Neuropathic Pain

John Manion[1], Thang Khuong[1], Leslie Caron[1], Greg Neely[1]

1. Charles Perkins Centre, The University of Sydney, Camperdown, NSW, Australia

Nerve injury can lead to devastating pain and for the majority of patients there are no effective therapies. Here we develop a model of neuropathic pain in *Drosophila*. Mechanistically and genetically, we prove for the first time that peripheral injury triggers a coordinated response from primary afferents resulting in loss of central inhibition, and culminating in a permanently sensitized pain circuit. While loss of central GABAergic tone was necessary for neuropathic pain to develop, disrupting inhibitory signaling was sufficient to trigger neuropathic sensitization without injury. To test the therapeutic potential of restoring central GABAergic inhibition, we generated and transplanted IPSC derived GABAergic inhibitory neurons into the spinal cord of mice suffering from neuropathic pain. Remarkably, iPSC-derived inhibitory neurons promoted lasting pain relief without side

effects. Together, these data highlight that central disinhibition is critical to neuropathic pain, and for the first time describe an iPSC transplant therapy that is an effective and long-lasting treatment for neuropathic pain.

212

When is a global repressor not a repressor? Genome-wide profiling of the Snail family transcription factor Worniu in neural stem cells

Caroline Delandre[1], Ciarán O'Mara[1], Owen J Marshall[1]

1. Menzies Institute for Medical Research, Hobart, TAS, Australia

The Snail family transcription factor Worniu is expressed specifically in neural stem cells (NSCs) within the fly brain, and has been previously shown to be required for larval NSC maintenance. In synthetic reporter assays the protein functions as a global repressor; however, its direct binding targets *in vivo* remain unknown.

Here, we describe the cell-type specific genome-wide binding of Worniu in larval NSCs, using our Targeted DamID technique combined with a novel deconvolution algorithm to dramatically increase the resolution. We relate these binding data to larval NSC transcriptional and chromatin state data, together with previously published transcriptional profiling data for both wild-type larval NSCs, and NSCs depleted for Worniu, to identify direct transcriptional targets. We find that Worniu binds to and represses a small number of genes involved in synaptic transmission, via a novel Trithorax-repressive chromatin state. Unexpectedly, however, we show that Worniu also binds promiscuously across open chromatin within the genome, including at active promoters and enhancers, inconsistent with a role as a global repressor. Combined, we suggest a model for Worniu's role in maintaining NSC identity within the brain.

213

Modelling transcriptional variability in single cell RNA-seq data during human embryogenesis captures changes in the regulation of critical developmental genes

Elizabeth A Mason[1], Shila Ghazanfar[2], Fredrik Lanner[3], Jean Yang[2], Christine A Wells[1]

1. University of Melbourne Centre for Stem Cell Systems, Parkville, VIC, Australia

2. School of Mathematics and Statistics, University of Sydney, Sydney, NSW, Australia

3. Department of Clinical Science, Intervention and Technology, Karolinska Institute, Stockholm, Sweden

Human development is a temporally and spatially ordered series of events that occur with remarkable precision. Embryogenesis appears predictable because we observe the average behaviour of many individual cells, even as the number of cells and transcriptional complexity increases during development. When we evaluate single molecules and transcripts the stochastic nature of gene expression is revealed, particularly in single cell RNA-seq experiments (scRNA-seq). Current methods reduce scRNA-seq data to a trajectory based on the abundance of key regulators of phenotype, and differential abundance is used to identify sub-populations. We present an alternative approach: measuring the transcriptional variability at the gene level informs the level of regulation imposed on it, reflecting an intrinsic property of development that is often overlooked. While linear models have successfully characterized the differences between phenotypes on average, they cannot account for stochastic differences captured by scRNA-seq experiments. Accurately determining abundance is further complicated by the sparseness of non-zero expression values. To address these challenges and evaluate gene expression during human pre-implantation embryogenesis, we applied a statistical mixture model to scRNA-seq data. Fitting the model on a gene-by-gene basis allowed us to evaluate shifts in the proportion of cells expressing a given gene (λ), and also the mean (μ) and standard deviation (σ) of expression. A correlation based analysis evaluated whether abundance (μ) and variability (σ) capture different aspects of transcriptional regulation. While each metric largely identified the same genes, the number and nature of relationships between them differed. Indeed, genes sharing correlated patterns of variability during development were enriched for motifs associated with developmental transcription factors. Variability was more effective than abundance at specifically detecting regulatory relationships during development, and with less redundancy. Our approach provides a gene-centric platform to evaluate population-based parameters of gene expression, while preserving the complexity of scRNA-seq data.

214

Coordinated up-regulation of splicing machinery components in a zebrafish intestinal development mutant harbouring a mutation in *gtf2h4*, a component of the general transcription factor complex, TFIIH

Elizabeth L Christie[1], Fansuo Geng[2], Stephen Mieruszynski[2], Karen Doggett[2], Janine Coates[2], Tanya A de Jong-Curtain[2], Christopher Love[3, 2], Oliver Sieber[2], Aliaksei Holik[2], Graham J Lieschke[4], Didier YR Stainier[5], Joan K Heath[2]

1. Ludwig Institute for Cancer Research, Parkville, VIC, Australia

2. Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia

3. Peter MacCallum Cancer Centre, Melbourne, Vic, Australia

4. Australian Regenerative Medicine Institute, Clayton, Vic, Australia

5. University of California, San Francisco, CA, USA

The transcription factor IIH (TFIIH) is a complex of ten proteins essential for the transcription of protein-coding genes and for DNA nucleotide excision repair (NER). Two of the core proteins, XPB/ERCC3 and XPD/ERCC2, exhibit 3′–5′ DNA helicase activity and contribute to the unwinding of DNA during RNA Pol2-mediated transcription and NER.

Through an ENU mutagenesis screen, we identified a zebrafish mutant, *sycorax*, with a single point mutation in *gtf2h4*, which encodes the TFIIH p52 protein that directly interacts with XPB. Loss of p52 expression during *sycorax* development results in a dysplastic intestinal epithelium comprising multi-layered, unpolarised cells, with a greater proportion of cells in S phase, compared to wild-type.

To study the impact of p52 depletion on transcription, RNA-seq was performed on individual, micro-dissected intestines from wild-type and *sycorax* larvae. Using locally developed software for differential expression and gene set enrichment analysis (Subread, edgeR, ROAST), we found that the intestine-specific transcriptome in *sycorax* contains 759 and 731 up-regulated and down-regulated genes, respectively. Interestingly, a significant upregulation of genes participating in mRNA splicing, transcription elongation and RNA processing components was identified, while genes that participate in metabolic and digestive functions were downregulated, indicative of impaired intestinal epithelial cell differentiation. These phenotypic and gene expression features are reminiscent of cancer cells.

As yet, the specific role of p52 has not been determined. Our zebrafish mutant provides an excellent *in vivo* model with which to study the function of p52 in the TFIIH complex during organogenesis, and to determine whether dysregulation of this role is relevant to cancer.

215

Identification of Hippo pathway target genes important for tissue growth.

Katrina A Mitchell[1, 2], Joseph H.A Vissers[1, 2], Jan Schroeder[3], Kieran F Harvey[1, 2]

1. Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia

2. Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, Victoria, Australia

3. Walter and Eliza Hall Institute, Melbourne, VIC, Australia

The evolutionarily conserved Hippo pathway plays an important role in organ size control by maintaining a precise balance of cell proliferation and cell death, and deregulation of Hippo signalling is observed in many cancer types. In *Drosophila melanogaster*, the downstream regulator of the pathway is the transcriptional coactivator protein Yorkie (Yki). Overactivity of Yki leads to hyperproliferation and overgrown tissues as a result of deregulation of target genes. Yki interacts with a number of transcription factors, for example Scalloped (Sd), to regulate the expression of genes. A repressor protein called Tondu-domain-containing growth inhibitor (Tgi) competes with Yki for Sd binding and the Sd-Tgi interaction results in repression of target genes. In this way, Yki promotes growth by relieving Sd and Tgi-mediated gene repression.

My phD research involves identifying the target genes regulated by the Hippo pathway in different growth contexts. Employing a powerful technique called targeted DamID, my research aims to identify and compare the target genes of Yki, Sd, and Tgi in wild type and hyperplastic tissue growth. Furthermore, in order to elucidate if Yki and Tgi require Sd for regulating their respective target genes, I will investigate Yki and Tgi target genes in a Sd mutant.

My studies have found that Yki, Sd, and Tgi bind to hundreds of genes in proliferating cells, including genes involved in signalling pathways such as MAPK signalling, Wnt signaling and the TGF-β pathway. Interestingly, many of these target genes are shared between Yki, Sd, and Tgi, indicating that they cooperate to control tissue growth by regulating the activity of many target genes and balancing target gene activation and repression.

Identifying the transcriptional targets of Hippo signalling will greatly expand our knowledge and understanding of how the Hippo pathway functions to control growth and how its deregulation can result in tumourogenesis.

216

Genetic and chemical approaches to inhibiting hyperplasia in a zebrafish model of hepatocellular carcinoma

Kimberly Morgan[1], Karen Doggett[1], Lachlan Whitehead[1], Stephen Mieruszynski[1], Brendon Monohan[1], Hendrik Falk[1], Ian Street[1], Joan Heath[1]

1. Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia

Liver cancer is the 3rd leading cause of cancer-related deaths worldwide, and represents a substantial and increasing health burden. As current therapeutic interventions offer limited survival improvements to patients, there is a critical need for the development of new and effective treatments. Multiple genes and molecular pathways, including RAS/MAPK signalling, are recurrently altered during oncogenic transformation. Epigenetic modifications, such as DNA and histone methylation, regulate many of these pathways fundamental to the genesis of cancer. Our study focuses on investigating the core oncogenic pathways in hepatocarcinogenesis.

Zebrafish provide a highly advantageous platform for in vivo drug discovery due to their small size, rapid maturation, optical transparency and high-throughput capabilities. We have utilised an inducible zebrafish hepatocellular carcinoma (HCC) model to identify and validate potential therapeutic targets. Upon the addition of doxycycline, an oncogenic version of *kras* fused to EGFP is expressed specifically in the liver. This results in hepatocyte hyperproliferation and liver enlargement. Through the application of 2-photon microscopy, the volume of normal and hyperplastic livers can be imaged in vivo and accurately quantitated.

Using this approach, we have shown that heterozygous mutations in genes participating in U12-dependent (minor class) splicing, ribosome biogenesis and nuclear pore formation significantly restrict $kras^{G12V}$-mediated liver overgrowth. Since these heterozygous mutations do not affect normal liver development, these experiments suggest that drugs designed to disrupt these processes may provide a therapeutic window that could be exploited clinically to restrict the growth of cancer cells without affecting surrounding healthy cells. Accordingly, we have utilised this model to evaluate the effectiveness of numerous drug treatments. Specifically, we have demonstrated that EPZ015666, an inhibitor of the protein arginine methyltransferase, PRMT5, has potent anti-proliferative activity, reducing liver hyperplasia. This study highlights the use of the zebrafish HCC model to identify and develop novel targeted therapies against cancer.

217

NEAT1 changes imatinib response and contributes to tyrosine kinase inhibitor-based resistance in a chronic myeloid leukemia cell line

Zeynep MUTLU[1, 2], Ruohan Li[2], Cigir Biray Avci[1], Archa Fox[2]

1. Department of Medical Biology, Medicine Faculty, Ege University, Turkey , The Scientific And Technological Research Council Of Turkey (TUBITAK), Ankara, Turkey

2. School of Human Sciences and School of Molecular Sciences, The University of Western Australia, Crawley, Australia

Chronic myeloid leukemia (CML) is a haematopoietic stem cell-based malignant disorder. CML is characterized by the presence of the philadelphia chromosome (9; 22), resulting in production of the BCR-ABL fusion protein with a tyrosine kinase activity. First line therapy for CML is Imatinib, a tyrosine kinase inhibitor(TKI) that targets BCR-ABL and improves prognosis for CML. However, the emergence of TKI resistance, largely due to mutation of the ABL kinase domain is a major problem.

Long non-coding RNAs (lncRNAs) are non-coding RNAs that are > 200 nucleotides in length and are involved in several diseases. Aberrant expression of the lncRNA Nuclear paraspeckle assembly transcript 1 (NEAT1) has been linked to many different cancer types. Two isoforms of NEAT1: NEAT1_1 (3.7 kb) and NEAT1_2 (23 kb) originate from the same promoter. In mammalian nuclei the expression of NEAT1_2 results in the formation of nuclear RNA-protein bodies named paraspeckles, wheras NEAT1_1 produces many smaller nuclear microspeckles. The function of NEAT1, microspeckles and paraspeckles still remains largely unknown in BCR-ABL mediated CML and TKI resistance mechanism.

In this study we aimed to evaluate the cytotoxic effect of imatinib in CML cells by altering NEAT1 expression levels and gene expression of NEAT1 targets which are associated with CML progression and TKI resistance mechanism. K562, a human chronic myeloid cell line, was treated with imatinib in a time and dose dependent

manner and cytotoxicity and NEAT1 levels were evaluated. In order to modulate NEAT1 expression levels, CRISPR-Cas9 genome editing was also performed in K562 cells.

Our findings showed that increased expression of NEAT1_1/microspeckles effects the imatinib response in K562 cells. One possibility we are investigating is that NEAT1 might target resistance-associated genes downstream of the BCR-ABL pathways that leads to imatinib resistance in CML. These results suggest that NEAT1 may be a therapeutic candidate in CML.

218

Whole genome dissection of the box jellyfish venom death pathway

Man Tat Lau[1], Greg Neely[1]

1. University of Sydney, Sydney, NSW, Australia

The box jellyfish *Chironex fleckeri* is the most venomous animal on the planet, with significant skin contact leading to extreme pain and death within minutes. Currently the only accepted emergency response to envenomation is continuous CPR to prevent death. Despite its potent actions on the nociception and cardiovascular systems, there is a lack of basic molecular insight into how this venom works. Here we perform the first ever molecular dissection of the jellyfish venom death pathway. We used genome-scale lenti-CRISPR mutagenesis with ~123,000 guide RNAs targeting all coding human genes and some noncoding RNA, screening for host components required for death after venom exposure. We identified hundreds of human genes that, when targeted, confer some resistance to box jellyfish venom. Among the host factors most enriched in our screen was the peripheral membrane protein ATP2B1, a calcium transporting ATPase implicated in the regulation of blood pressure and a likely candidate for the venom receptor. Jellyfish venom resistance in the absence of ATP2B1 was confirmed in multiple cell types and this protection was long lasting. Analysis of venom resistance genes highlighted numerous new pathways not previously implicated in cell death, and we have further validated many of these factors are required for box jellyfish-mediated death. Guided by these data, we have tested multiple compounds that target the box jellyfish death pathway, and have identified a new “anti-venom” drug that can completely suppress cell death even when added up to 15 minutes after venom exposure. This compound is safe for human use, stable at room temperature, and may represent a new therapy for box jellyfish and potentially other venoms. These results highlight the power of whole genome CRISPR screening to investigate unknown molecular mechanisms and new biology.

219

The Tumor Suppressor EYA4 Functions in DNA Double Strand Break Repair

Christopher B Nelson[1], Alexander P Sobinoff[1], Christopher G Tomlinson[1], Eloise Dray[2], Hilda A Pickett[1]

1. Telomere Length Regulation Unit, Children's Medical Research Institute, Westmead, NSW, Australia

2. Translational Research Institute, Queensland University of Technology, Brisbane City, QLD, Australia

The transcription factor and atypical protein phosphatase EYA4 (a member of the Eyes Absent Protein Family) has emerged in recent years as an important tumor suppressor. To date, mutation or downregulation of EYA4 has been associated with a range of tumor types including esophageal, lymphoma, uterine, colon, and oral. However, little is known about the biology of EYA4 in the context of carcinogenesis. Interestingly, other members of the Eyes Absent protein family have been implicated in double strand break (DSB) repair, and recently EYA4 overexpression was shown to promote gamma-H2AX levels and retention following ionizing radiation (IR), suggesting that EYA4 may also function in DSB repair.

We have found that EYA4 knock down (KD) leads to a decrease in sister chromatid exchanges (SCEs) and an increase in chromosome fusions, which result from homologous recombination and end joining activities, respectively. Additionally, DNA fragmentation is decreased in EYA4 KD cells compared to controls early after ionizing radiation (IR), as measured by the neutral comet assay. This observation is consistent with a role for EYA4 in the repression of end joining. Despite this, EYA4 may not act locally at DSBs, as immunofluorescence has revealed that it does not form discrete foci at them. Finally, phosphorylation of global PIKK kinase substrates is enhanced by EYA4 KD post IR, suggesting that the phosphatase activity of EYA4 likely contributes to its role in DSB repair. Further work will seek to identify specific targets of EYA4 phosphatase activity and how these events contribute to DSB repair. To identify potential candidates, we are currently employing a BIO-ID proteomics screen for EYA4 interacting partners with and without treatment with IR.

220

Topoisomerase 2A is essential for maintenance of mitotic chromosome structure

Christian Friberg F Nielsen[1], Tao Zhang[1], Paul Kalitsis[1], Damien F Hudson[1]

1. Murdoch Childrens Research Institute, Parkville, VICTORIA, Australia

Faithful propagation of genomic material during the cell division cycle is crucial for the cells of our body to keep malignancy in check and ageing at bay. Cells prepare our DNA for division by packing it tightly into highly condensed and resolved chromosomes. Chromosomes that are not correctly folded cause major problems in mitosis and lead to genome instability through bridging or chromosome loss during anaphase. Maintaining chromosomes in a compacted disentangled state before division is therefore essential, yet, the process of chromosome maintenance is poorly understood. Topoisomerase 2A (TOP2A), a DNA metabolising enzyme commonly targeted by chemotherapy for cancer, is known to be involved in shaping and disentanglement of chromosomes. Previous studies have suggested a role in the establishment of mitotic chromosome structure. A major difficulty with studying TOP2A in mitosis is that it has key functions in transcription and DNA replication. Therefore systems such as RNAi that deplete protein over multiple cell cycles make it difficult to define primary phenotypes. To precisely pinpoint the function of TOP2A in mitosis we have engineered a human cell line that is capable of ultra-rapid conditional TOP2A depletion using the novel auxin/degron system. TOP2A depletion causes chromosome bridging and cells arrest before cytokinesis. Within 30 minutes of depletion of TOP2A, chromosomes in mitosis begin to untangle and lose structure. Correspondingly, when TOP2A is rapidly removed from cells arrested in prometaphase overnight with hypercondensed chromosomes, the chromosomes unravel and appear to twist and re-entangle. Our novel data argue that TOP2A has a key role in the maintenance of mitotic chromosome compaction and suggest paradoxically that the enzyme is a key structural component of mitotic chromosomes.

221

Investigating the evolution of truffle-like fungi

Andy R Nilsen[1], Kyrin R Hanning[1], Gareth Gillard[2], David A Orlovich[1], Tina C Summerfield[1], Christopher M Brown[1]

1. University of Otago, Dunedin, Otago, New Zealand

2. Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway

Truffle-like fungi have evolved from mushroom-like ancestors and are unable to forcibly discharge their spores, rather sequestering them in closed caps. The truffle-like morphology has independently evolved on multiple occasions throughout the fungal kingdom. It has been suggested to have arisen in response to dry climates where fruit body desiccation prior to maturation would severely reduce reproductive success. Evidence from time of emergence studies of truffle-like fungi indicates that past aridification events in Australia may have played an important role in their evolution. The inability to forcibly discharge spores also makes truffle-like fungi highly dependent on animal vectors for dispersal. To encourage dispersal by animals Truffle-like fungi have evolved different attractants, most notably the production of aromatic compounds and bright colours.

While evolutionary drivers have been suggested for the selection of truffle-like fungi, very little is understood about the underlying molecular mechanisms that have facilitated the transition from the mushroom-like state to the truffle-like habit.

Here we present current knowledge in the evolution of truffle-like fungi with an emphasis on New Zealand taxa.

222

Mutational Landscape of Familial Breast Cancers

Katia Nones[1], Julie Johnson[2], Xavier De Luca[2], Sriganesh Srihari[3], Kaltin Ferguson[2], Lynne Reid[2], Amy M McCart Reed[2], Stephen Kazakoff[1], Ann-Marie Patch[1], Felicity Newell[1], Vanessa Lakis[1], Oliver Holmes[1], Qinying Xu[1], Scott Wood[1], Conrad Leonard[1], The Kathleen Cuningham Foundation Consortium for Research into Familial Aspects of Breast Cancer (kConFab)[4], Australian Breast Cancer Tissue Bank (ABCTB)[5], Brisbane Breast Bank (BBB)[2], Mark Ragan[3], Georgia Chenevix-Trench[1], Kum Kum Khanna[1], Sunil Lakhani[2], John V Pearson[1], Nic Waddell[1], Peter Simpson[2]

1. QIMR Berghofer, Brisbane, QLD, Australia

2. UQ Centre For Clinical Research, The University Of Queensland, Brisbane, Queensland, Australia

3. Institute of Molecular Biosciences, The University Of Queensland, Brisbane, QLD, Australia

4. The Kathleen Cuningham Foundation Consortium for Research into Familial Aspects of Breast Cancer , Peter MacCallum Cancer Centre, Melbourne, VIC, Australia

5. Australian Breast Cancer Tissue Bank , University of Sydney at the Westmead Institute for Medical Research, Sydney, NSW, Australia

Mutations in BRCA1 and BRCA2 genes are associated with approximately 45% of families with multiple cases of breast cancer and up to 90% of families with both breast and ovarian cancer. However, a proportion of women with hereditary breast cancer have no identified pathogenic variants in genes predisposing them to breast cancer. To better understand the aetiology of familial breast cancer we are conducting a whole genome analysis of tumours from familial cases with *BRCA1 (n=26)*, *BRCA2 (n=22)* and non-*BRCA1/2 (n=32)* germline mutation carriers. Biospecimens and clinical data were obtained from kConFab, the Brisbane Breast Bank and the Australian Breast Cancer Tissue Bank. Whole genome sequencing (WGS) was performed for 80 matched tumour/normal pairs (60x and 30x coverage, respectively). Whole genome methylation arrays were analysed for 67 tumours and RNASeq for 37 tumours. Comparative analysis with data from 99 sporadic breast tumours from The Cancer Genome Atlas is underway to identify potential similarities or differences between familial and sporadic tumours. The complete spectrum of somatic and germline of mutations has been evaluated, including SNPs, indels, copy number changes and structural rearrangements, and mutational signatures. Our aim is to identify potential mutations or mechanisms underlying familial non-*BRCA1/2* breast cancers. *BRCA1*, *BRCA2* and non-*BRCA1/2* tumours exhibited a different burden of mutations, different spectrum of mutational signatures and different telomere length. In 5 non-*BRCA1/2* cases mutational signatures suggests that deficiency in the Homologous Recombination (HR) pathway might be driving the development of those tumours but not through *BRCA1* or *BRCA2* mutation. We will present our most recent findings in detecting the mechanisms underlying familial breast cancers that in the future could help the management of families where the underlying cause is unclear.

223

Single-cell RNA profiling reveals developmental lineage relationships between the mouse mammary epithelial cells

Bhupinder Pal[1], Yunshun Chen[1], François Vaillant[1], Paul Jamieson[1], Lavinia Gordon[2], Melissa Davis[1], Anne C Rios[1], Kevin Liu[1], Felicity Jackling[1], Naiyang Fu[1], Stephen Wilcox[1], Geoffrey J Lindeman[1], Gordon K Smyth[1], Jane E Visvader[1]

1. Walter and Eliza Hall Institute of Medical Research, Parkville, Vic, Australia

2. Australian Genome Research Facility, Parkville, VIC, Australia

Here we report single-cell RNA profiling of mouse mammary epithelial cells across four developmental stages in the post-natal gland. Notably, the epithelium undergoes a large-scale shift in gene expression from a relatively homogeneous basal-like program in pre-puberty to distinct lineage-restricted programs in puberty. Interrogation of single-cell transcriptomes reveals different levels of diversity within the luminal and basal compartments, and identifies an early progenitor subset marked by CD55. Moreover, we uncover a luminal transit population and a rare mixed-lineage cluster amongst basal cells in the adult mammary gland. Together these findings point to a developmental hierarchy in which a basal-like gene expression program prevails in the early post-natal gland prior to the specification of distinct lineage signatures, and the presence of cellular intermediates that may serve as transit or lineage-primed cells.

224

Genomics of malignant pleural mesothelioma

Ann-Marie Patch[1], Katia Nones[1], Venkateswar Addala[1], Felicity Newell[1], Stephen Kazakoff[1], Sophie Sneddon[2], Ian Dick[2], Ebony Rouse[2], Justine Leon[2], John V Pearson[1], Dean Fennel[3], Jenette Creaney[2], Bruce Robinson[2], Nic Waddell[1]

1. Genetics & Computational Biology, QIMR Berghofer Medical Research Institute, Herston, QLD, Australia

2. National Centre for Asbestos Related Disease, School of Medicine and Pharmacology, University of Western Australia, Nedlands, WA, Australia

3. Cancer Research UK Centre Leicester, University of Leicester & University Hospitals of Leicester NHS Trust, Leicester, United Kingdom

Malignant pleural mesothelioma (MPM) is an incurable and aggressive cancer that occurs on pleural membranes covering the lungs and is associated with exposure to asbestos fibres. There are >15,000 cases per annum worldwide. Regulation against asbestos has slowed the increase rate of Australian deaths however for some Asian countries, continued unregulated use of asbestos means that their disease burden is set to

increase. Combinations of surgery, chemotherapy and radiotherapy all feature in the clinical management but the overall survival remains at 9-24 months. Therefore new therapeutic approaches are needed.

We are assembling the world's largest MPM whole genome sequencing resource, comprising sequencing for 54 patients with accompanying methylation and transcriptome sequencing where possible. Our results have confirmed driver gene mutations for CDKN2A, BAP1, *NF2*, *SETD2* and *TP53*. For CDKN2A, BAP1 and NF2 the mutation spectrum commonly includes disruption by deletion and rearrangements breakpoints, indels and/or truncating point mutations.

Intra-tumoral heterogeneity can affect response to all types of treatment and may be identified from the whole genome sequencing. We have multiple samples from 5 MPM patients that allow us to explore inter sample evolutionary changes and relate these to the clinical course of treatment. MPM samples often have a low number of branch mutations and temporally separated samples appear largely stable.

Immunotherapy is an emerging cancer therapeutic approach with some exceptional responses in metastatic, recurrent and heavily mutated cancers. MPM cancers generally have a low mutation burden, in our cohort it is only 1.23 mutations per mega base, however recent clinical trials suggest that MPM patients may respond to immunotherapy. We have employed genomic HLA typing, neoantigen prediction and immune environment profiling to characterise the immune setting for MPM patients. Our project will characterise and integrate genomics, immune environment and clinical data to identify the underlying genomic mechanisms for therapeutic success.

225

Human *SRY* regulates haploid specific Protamine (*PRM1*) promoter activity

Deepali Pathak[1], Arka Baksi[1], S S Vasan[2], Rajan R Dighe[1]

1. Indian Institute of Science, Bangalore, KARNATAKA, India

2. Manipal Ankur Fertility Clinic, Bangalore, Karnataka, India

Publish consent withheld

226

Predicting the outcome of breast cancer using novel RNA-Seq analysis

Andrew Pattison[1], Paul Harrison[2, 1], Traude Beilharz[1]

1. Department of Biochemistry and Molecular Biology, Monash University, Melbourne, VIC, Australia

2. Monash Bioinformatics Platform, Monash University, Melbourne, VIC, Australia

With the exception of skin cancers, breast cancer is the most common cancer affecting women. While progress has been made in the detection of primary breast tumours, there are few genomic tests that are able to accurately predict outcome. Current genomic tests such as as Mammaprint and Oncotype DX are not widely available and are only suitable for early stage tumours, with additional restrictions applying depending on the test used. We sought to derive a new predictor of breast cancer outcome from TCGA RNA-Seq data that can provide an accurate indication of prognosis, even in later stage tumours. Alternative polyadenylation (APA) is the process whereby the poly (A) tail is added to the 3' untranslated region (3' UTR) of a messenger RNA (mRNA) at one of multiple possible sites, changing 3' UTR length and potentially the regulatory elements that bind to it. APA has been suggested to be predictive of tumour outcome and can be inferred from RNA-Seq data. We used elastic net linear modelling to select coefficients that best predict relapse free survival from clinical, APA and gene expression data. The best model was generated using a combination of all 3 data types, with common clinical indicators playing only a small role. When validated using 10 fold cross validation, patients with a score higher than the median generated by our model were at least 3 times less likely to die of cancer than those with a score below the median ($p << 0.01$). Our ultimate aim is to derive an accurate genomic test for breast cancer outcome that can be applied to all breast tumours and is less reliant on clinical data. This test could potentially be implemented using the in house M-PAT approach, for substantially less than the cost of a full RNA-Seq experiment.

227

Computational modeling reveals 4D genome reorganization during stem cell differentiation

Jonas Paulsen[1], Philippe Collas[1]

1. University of Oslo, Oslo, OSLO, Norway

We have developed and published a user-friendly computational 3D modeling tool that simulates positions of topologically-associated domains (TADs) relative to each other and to the nuclear periphery. The

computational tool, called Chrom3D, integrates chromosome conformation capture (Hi-C) and lamin-associated domain (LAD) datasets to generate structure ensembles that recapitulate radial distributions of TADs detected in single cells. We use our tool to study the dynamic structural 4D reorganization of TADs during stem cell differentiation, using Hi-C and lamin ChIP-seq in multiple time steps. The resulting 4D models reveal TAD-TAD interactions involved in large scale dynamic alterations in chromatin architecture linked to gene expression changes across time points.

228

A Functional Genomic Approach to Identifying New Motor Neuron Disease Genes and Drug Targets

Jodie Pearlman[1], John Manion[1], Rachel Walters[1], Harry Towler[1], Yong Lin[1], Thang Khuong[1], Greg Neely[1]

1. Charles Perkins Centre, The University of Sydney, Sydney, NSW, Australia

It has been suggested that defective synaptic transmission at the neuromuscular junction (NMJ) underlies the aetiology of motor neuron disease. However, our understanding of the regulation of synaptic transmission remains incomplete. In order to investigate this we identified 420 candidate synaptic regulators by combining MND genome wide association studies, gene enrichment at neuropil and hits from genome wide studies of pain and lethality in *Drosophila*. Using RNAi we specifically knocked down these genes in *Drosophila* motor neurons to identify new synaptic regulators through climbing behaviour. In parallel, we are investigating genes that modulate a MND phenotype (expression of TDP-43 in eye neurons). So far we have identified around 100 genes that have a role in motor neuron function. Gene ontology analysis of these hits suggests significant associations with microtubule organisation. We will now complete our screens and further examine candidate genes with synaptic assays including morphology and electrophysiology.

229

A general framework for evaluating cross-platform concordance in genomic studies

Timothy J Peters[1], Hugh J French[2], Stephen T Bradford[1], Ruth Pidsley[1], Susan J Clark[1], Terry Speed[3]

1. Garvan Institute of Medical Research, Darlinghurst, NSW, Australia

2. South Western Sydney Clinical School, Faculty of Medicine, University of New South Wales, Liverpool, NSW, Australia

3. Bioinformatics Division, Walter & Eliza Hall Institute, Parkville, VIC, Australia

The reproducibility of scientific results from multiple sources is critical to the establishment of scientific doctrine. However, when characterising various genomic features (transcript/gene abundances, methylation levels, allele frequencies and the like), all measurements from any given technology are estimates and thus will retain some degree of error. Hence defining a “gold standard” process is dangerous, since all subsequent measurement comparisons will be biased towards that standard.

In the absence of a “gold standard” we instead empirically assess the precision and sensitivity of a large suite of genomic technologies via a consensus modelling method called the row-linear model. This method is an application of the American Society for Testing and Materials (ASTM) Standard E691 for assessing interlaboratory precision and sources of variability across multiple testing sites. We analyse three datasets (two RNA expression, one DNA methylation), each containing both sequencing and array technologies, allowing a direct per-technology, per-locus comparison of sensitivity and precision across all common loci. We assess the performance of a number of technologies including the Infinium MethylationEPIC BeadChip, whole genome bisulfite sequencing (WGBS), two different RNA-Seq protocols (PolyA+ and Ribo-Zero) and five different gene expression array platforms.

We implement and showcase a number of applications of the row-linear model, including direct comparisons of the sensitivity and precision of these platforms, correlation with known interfering traits related to probe and target biochemistry such as GC content, CDS length and cross-hybridisation, and the effect of normalisation on DNA methylation arrays. Our findings demonstrate the utility of the row-linear model in evincing varying levels of concordance between measurements on these platforms, serving as a process for identifying reproducibility caveats in studies where cross-platform validation is performed.

230

Developing CRISPR-Cas9 Based Gene Drives in *Mus musculus*

Chandran Pfitzner[1, 2], Fatwa Adikusuma[1, 2], Melissa White[1, 2, 3], Sandra Piltz[1, 2, 3], James N Hughes[1], Paul Q Thomas[1, 2, 3]

1. Molecular & Cellular Biology, University of Adelaide, Adelaide, SA, Australia

2. South Australian Health and Medical Research Institute, Adelaide, SA, Australia

3. SA Genome Editing Facility, Adelaide, SA, Australia

Invasive vertebrate pests including mice cause significant environmental damage and agricultural loss of productivity. Current control and eradication methods have limited efficacy. A potential solution to this problem is the genetic modification of entire populations of wild mice to suppress their numbers. A CRISPR-Cas9 based gene drive is a genetic construct that promotes its own inheritance and can therefore spread through a given population. This is achieved by incorporating the components of the CRISPR-Cas9 system into the mouse genome where it can be inherited from one parent and subsequently copy itself to the wild type chromosome inherited from the other parent. This approach has the potential to eradicate invasive mice populations through addition of "cargo" genes that induce phenotypes such as sex-reversal, sterility, or embryonic non-viability leading to a population crash. CRISPR-Cas9 gene drives have so far only been shown to work effectively in a small number of species including flies, mosquitoes, and yeast. We have constructed a “split” gene drive in mice driven by a ubiquitous CMV promoter. Our data show that the gene drive system is able to effectively cut the wild type allele but is unable to copy itself, instead generating indels. Work is currently underway to produce a gene drive under the control of the germline-expressing *VASA* promoter which offers greater promise at effectiveness. This is the first attempt at developing CRISPR-Cas9 gene drive technology in vertebrates and provides an important first step towards assessing their potential for population suppression of invasive mice.

231

Insights into the biogenesis and possible functions of exonic circular RNA

Chikako Ragan, Nikolay Shirokikh, Greg Goodall, Thomas Preiss

Circular RNAs (circRNAs) exhibit unique properties due to their covalently closed nature. Models of circRNAs synthesis and function are emerging, with some known to be translated or act as microRNA sponges, but much about the biogenesis and functions of this surprisingly prevalent class of RNA remains undefined. We identify exonic circRNAs from human and mouse RNA-sequencing datasets, documenting multiple new examples. Applying stringent criteria, we find miRNA binding sites only in a minority of circRNAs and potent ‘sponging’ of a single microRNA by a single circRNA is scarce. Nevertheless, we detect potential for groups of circRNAs acting as ‘distributed’ microRNA sponges. Multiple circRNAs co-sediment with ribosomes, indicative of translation potential. Consistently, we find these circRNAs to be enriched for short open reading frames started by CUG and GUG codons. CircRNAs typically span few exons and their host genes show a pronounced preference for early back-spliced acceptor position and long flanking introns. Analysis of RNA polymerase II (Pol II) occupancy reveals distinctive transcription dynamics. Overall, Pol II traverses circRNA producing regions at accelerated speeds, but shows pronounced pausing over the back-splicing acceptor site. These features likely represent kinetic preconditions for circRNA production that are augmented by looping interactions between flanking introns.

232

The identification of rare variants in Tasmanian prostate cancer pedigrees using whole-genome sequencing

Kelsie Raspin[1], Liesel M FitzGerald[1], Marthick R James[1], Matt A Field[2], Banks Annette[1], Donovan Shaun[3], Thomson J Russell[4], Roslyn Malley[1, 5], Joanne L Dickinson[1]

1. Menzies Institute for Medical Research - University of Tasmania, Hobart, TASMANIA, Australia

2. James Cook University, Smithfield, QUEENSLAND, Australia

3. Hobart Pathology, Hobart, TASMANIA, Australia

4. University of Western Sydney, Sydney, NEW SOUTH WALES, Australia

5. Royal Hobart Hospital, Hobart, TASMANIA, Australia

Prostate cancer (PCa) is a highly heterogeneous disease with studies suggesting a degree of heritability greater than other cancers. It has been found that as much as 58% of disease risk can be explained by heritable factors. While more than 150 common genetic risk variants have been identified, these variants still only explain a minor portion of risk, are largely of low to moderate effect size, and are functionally ambiguous.

There has recently been significant success in the discovery of rare genetic variants contributing to complex disease, through next-generation sequencing of large families, where rare variants are enriched and there is reduced genetic complexity. Here, we have applied whole-genome sequencing (WGS) to several large Tasmanian PCa pedigrees with the aim of identifying rare genetic variants contributing to the development of PCa.

Thirty-seven individuals from six PCa pedigrees were WGS on the Illumina HiSeq X™ Ten platform. Variants were prioritised on a per-family basis by frequency (<1% in 1KGP, UK10K, ExAC and ESP), segregation with disease, mutation type (missense, nonsense or splice) and predicted functional consequence (CADD, PolyPhen and SIFT). Unaffected older male relatives and population controls were also WGS and used to prioritise variants based on non-sharing. After additional genotyping in our larger familial and population datasets, familial-based association testing found two variants to be significantly associated with PCa risk. A rare variant in the non-coding region of the EZH2 gene was identified in two Tasmanian families, and has the potential to effect splicing. *EZH2* is constitutively overexpressed in many cancer, including PCa, and therefore, functional studies are currently underway to determine the effect of the mutation on gene and protein expression.

This study was designed to identify new biological pathways involved in the pathogenesis of familial PCa and may lead to novel diagnostic and therapeutic targets for PCa patients.

233

Splicing factors as novel regulators of oncogenic miRNAs

Madara Ratnadiwakara[1], Charlotte Nejad[2], Michael P Gantier[2], Traude H Beilharz[1], Helen Abud[1], Minna-Liisa Anko[1]

1. Monash University, Clayton, VIC, Australia

2. Hudson Institute of Medical Research, Clayton, VIC

MicroRNAs (miRNAs) regulate gene expression post-transcriptionally by fine-tuning mRNA levels and translation. miRNA biogenesis is tightly regulated to maintain specific miRNA expression patterns in different tissues and developmental stages, misexpression leading to pathological conditions. miRNAs are transcribed as longer precursors that undergo multiple processing steps before the mature miRNAs reach their target mRNAs in the cytoplasm. Although the RNA binding proteins (RBPs) Drosha, DGCR8 and Dicer are the essential components of the miRNA processing pathway, other RBPs, have recently been identified as critical regulators of miRNA biogenesis. The CNNC motif bound by the SR protein splicing factor SRSF3 has been shown to specify pri-miRNA hairpins and *in vitro*studies have demonstrated SRSF3 enhancing pri-miRNA processing in a CNNC dependant manner. However, the *in vivo* functional relevance of SRSF3 activity in miRNA processing has remained unexplored. Our analysis of SRSF3 binding sites in embryonic stem cells (ESCs) identified SRSF3 binding at the CNNC motif particularly in miRNAs located in polycistronic miRNA clusters. We demonstrate that SRSF3 depletion in ESCs leads to reduced levels of mature miRNAs without affecting the levels of the pri-miRNAs. Intriguingly, the processing of specific miRNAs within miRNA clusters is individually regulated, leading to differential expression of mature miRNAs derived from the same pri-miRNA. Furthermore, SRSF3 is frequently overexpressed in tumour cells and misexpression of miRNAs is a characteristic of many cancers. Our analysis of SRSF3, pri-miRNA, miRNA and miRNA-target expression in human tumour-normal pairs demonstrates that SRSF3 may confer some of its pathological properties through the control of miRNAs. Thus, our work reveals a novel mechanism regulating the hallmark properties of cancer cells and highlights the roles of multifunctional RBPs in gene regulation in health and disease.

234

Expanding the RNA-guided endonuclease toolkit for mouse genome editing.

Louise Robertson[1], Daniel Pederick[1], Sandra Piltz[2, 1], Melissa White[2, 1], Marie Ahladas[3], Fatwa Adikusuma[1], Paul Thomas[3, 2, 1]

1. Department of Molecular and Cellular Biology and Robinson Research Institute, University of Adelaide, Adelaide, SA, Australia

2. SA Genome Editing, Adelaide, SA, Australia

3. South Australia Health and Medical Research Institute , Adelaide, SA, Australia

The RNA-guided endonuclease CRISPR/Cas system from *Streptococcus pyogenes* (SpCas9) has been widely used for the generation of genetically modified mice via zygote micro-injection. Although double-strand breaks (DSBs) are induced efficiently by SpCas9, it requires an NGG proto-spacer adjacent motif (PAM) at the target site. This PAM requirement is particularly limiting when performing targeted DNA insertions/substitutions where the DSB must be in close proximity to the mutation site. Recent development in

the CRISPR/Cas field has seen the characterisation of alternative RNA-guided endonucleases that require different PAM sequences. Here we investigate whether these different CRISPR/Cas systems can also efficiently induce targeted mutations when injected into mouse zygotes. We tested the following endonucleases: SpCas9 VQR (NGAN PAM), SpCas9 VRER (NGCG PAM), AsCpf1 (TTTN PAM) and SaCas9 KKH (NNNRRT PAM), as well as the wild type (WT) SpCas9 (NGG PAM). As expected, WT SpCas9 efficiently induced DSBs with 95% of embryos containing mutations. AsCpf1 (30% mutant embryos), VQR (11% mutant embryos) and VRER (6% mutant embryos) also generated mutations but with lower efficiency. Interestingly, the SaCas9 KKH variant generated mutations in 56% of embryos and 93% of these had remaining wildtype alleles, indicating heterozygosity. We tested this property of SaCas9 KKH by injecting it alongside a ssDNA template and observed that when oligo insertion occurred, the other allele always remained unmodified. Conversely, WT SpCas9 mutated the second allele frequently (50%), a problem when targeting a gene that causes nullizygous embryonic lethality. Our findings extend the toolbox of RNA-guided endonucleases for mouse genome editing, not only broadening the PAM recognition options, but also presenting SaCas9 KKH as an attractive alternative for targeting genes essential for embryonic survival.

235

The A/HeJ Mouse: Dysfunction in Sex Development

Jessica Robinson[2, 1], Alison Graham[2], Paul Kalitsis[2, 1]

1. Department of Paediatrics, The University of Melbourne, Parkville, Victoria, Australia

2. Murdoch Children's Research Institute, Parkville, Victoria, Australia

The development pathway of the male gonad, the testis, occurs in response to the expression of the Y chromosome-linked *Sry* gene within the somatic cells of the primordial gonad. In mice, the expression of *Sry* is tightly spatially and temporally regulated, with testis development only inducible within a brief time window during embryonic development. Disruption or delay of *Sry* expression can result in complete XY sex reversal, giving rise an ovary, or an intermediate gonadal phenotype, producing an ovotestis. A mouse strain with a reported anomaly in male sex development is the A/HeJ mouse, first reported in 2008 by Hunt et al in “The mouse A/HeJ Y chromosome: Another good Y gone bad”. This mouse strain exhibits a disturbance in the male sex determination pathway, with 4% of non-productive males presenting as overt hermaphrodites, and a further 17% having small testes with no epididymal sperm. It was hypothesised that there is a deletion at or near the Y chromosome centromere which affects the expression of *Sry* during embryogenesis. This study examines the region surrounding the Y chromosome centromere of A/HeJ and other closely related strains. Epigenetic and gene expression analyses have been performed on *Sry* in this strain to determine the aetiology of this sex development disturbance.

1. Hunt PA; Jackson JM; Horan S; Lawson CA; Grindell L; Washburn LL; Eicher EM. 2008. The mouse A/HeJ Y chromosome: Another good Y gone bad. Chromosome Res 16(4):623-36. PubMed: 18483871MGI: J:136868

236

MASTL overexpression promotes chromosome instability and metastasis in breast cancer

Samuel Rogers[1], Rachael McCloy, Benjamin Parker, David Gallego-Ortega, Andrew Law, Venessa Chin, James Conway, Dirk Fey, Medina Krcic, niantao deng, ewan millar, Sandra O'Toole, Alex Swarbrick, Paul Timpson, Elizabeth Caldon, David Croucher, david james, Neil Watkins, Andrew Burgess

1. Children's Medical Research Institute, Westmead, NSW, Australia

Chromosome instability is the most common genomic aberration, implicated in both cancer development, and malignant progression. Despite driving multi-drug resistance and metastasis; a details of the molecular origin of chromosome instability are still lacking. What is known, is that chromosome instability arises from aberrant mitotic division. Previous work by our lab identified that mitosis is under the control of MASTL kinase, an inhibitor of PP2A phosphatase. Loss of MASTL is embryonically lethal, while perturbation of its activity causes chromosome segregation errors, mitotic collapse, and cytokinesis failure. Therefore, we investigated whether MASTL has a role in the origin of chromosome instability, and thus oncogenic transformation. Not surprisingly, MASTL is commonly amplified or overexpressed in cancer. We show that high MASTL expression correlates with increased nuclear pleomorphisms in breast cancer, and poor patient survival in breast, ovarian and lung cancer. Overexpression of MASTL at biologically relevant levels in normal MCF10A breast cells delayed cell-cycle transit time, and increased the rate of chromatin bridges, suggesting that MASTL is directly involved in chromosome instability. These genomic aberrations were accompanied by a relief of contact inhibition and the induction of a partial epithelial-to-mesenchymal transition which are characteristics of oncogenic

transformation. To determine the mechanism for MASTLs role in transformation we undertook a global phosphoproteomic analysis. This study identified that key members of desmosome assembly, stress-kinase signalling, and the actin-cytoskeleton were deregulated by MASTL overexpression. Importantly, knockdown of MASTL prevented invasion and metastasis of breast cancer cells in *in vitro,* organotypic, and intraductal mouse models. Taken together these results suggest that MASTL overexpression contributes to chromosome instability and metastasis and defines a novel role for MASTL as a regulator of epithelial to mesenchymal transition. This study provides a mechanistic basis for the action of MASTL in breast cancer, and its negative effects on survival and metastasis.

237

Overcoming off target effects in CRISPR and RNAi loss of function screens

Joseph Rosenbluh[1]

1. Biochemistry and Molecular Biology, Monash University, Clayton, VIC, Australia

Loss of function screens provide the means to modulate gene expression and permit to systematically study gene function. By comparing RNAi and CRISPR we found that seed mediated RNAi off target effects complicate the interpretation of RNAi based approaches and that CRISPR provides a highly specific approach to modulate gene expression. However, DNA amplifications which are common in many cancer cell lines mediate a Cas9 induced DNA damage response that results in a gene independent proliferation arrest. To overcome this hurdle, we have adopted CRISPR interference (CRISPRi) as an additional approach to suppress gene expression. We found that CRISPRi is not effected by DNA amplifications however, bidirectional promoters (present in 10% of human genes) score as false positives in CRISPRi screens. We demonstrate the utility of combining these approaches as a highly specific approach to study how genes are regulated in normal cells and how they are deregulated during disease progression.

238

Characterising Human Variants of the Transcription Factors NPAS3 and NPAS4 Identified in Patients with Mental Illness

Joseph J Rossi[1], David C Bersten [1], Murray L Whitelaw[1]

1. Department of Molecular and Cellular Biology, University of Adelaide , Adelaide , SA, Australia

The basic-Helix-Loop-Helix PER-ARNT-SIM (bHLH-PAS) transcription factors Neuronal PAS 3 and 4 (NPAS3/4) are expressed in several structures of the mammalian brain. To influence transcription, NPAS3 and NPAS4 must heterodimerise with their partner protein Aryl Hydrocarbon Receptor Nuclear Translocator (ARNT). NPAS3 is important for neuronal proliferation during development and in the dentate gyrus of the adult hippocampus. In contrast, NPAS4 is expressed primarily in the adult brain where it maintains neural circuit homeostasis via synaptic modulation. Despite these differences, studies using mouse models have implicated both NPAS3 and NPAS4 in schizophrenia and intellectual disability. However, there is currently a significant gap in the literature concerning mutations in the human NPAS3/4 genes in association with mental illness. There have only been two mutations identified for the human NPAS3 gene, whilst no mutations have been identified for NPAS4. Recently, we have utilised unpublished exome sequencing data to identify individuals with mental illness that have either non-synonymous or truncating mutations for NPAS3/4. It is necessary to identify mutants with deficient transcriptional activity to determine if they are contributing to the disease phenotype. To identify deficient mutants, luciferase reporter gene assays were utilised and only the truncating mutations of NPAS3 and NPAS4 had decreased transcriptional activity. To then investigate a possible mechanism for the reduced reporter activity, we determined the ability of the NPAS3/4 proteins to heterodimerise with their obligate partner ARNT via co-immunoprecipitation. We plan on performing additional mechanistic experiments assessing protein localisation, protein turnover and DNA binding.

239

Modulation of higher order chromatin structures associated with virus-specific killer T cell differentiation

Brendan Russ[1], Moshe Olshanksy[1], Zhoaren Ye[2], Jasmine Li[1], Sara Tomei[1], Phillipe Collas[3], Cornelius Murre[2], Stephen Turner[4, 1]

1. Microbiology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria, Australia

2. Division of Biological Sciences, University of California, San Diego, San Diego, CA, USA

3. Institute for Basic Medical Sciences, University of Oslo, Oslo, Norway

4. Monash University, Clayton, VIC, Australia

A cardinal feature of protective T cell immunity is that activation induces a program of proliferation and differentiation resulting in the acquisition and long term maintenance of lineage-specific function. While it is known that the different functional capacities of naïve, effector and memory T cells are underscored by unique transcriptomes, how these arise and are maintained is poorly understood. We have previously demonstrated that differentiation-specific transcriptional signatures are established through modulation of histone post translational modifications (PTMS) and binding of activating and repressive transcription factors (TFs) to regulatory sequences including promoters and transcriptional enhancers (TEs)[1, 2]. However, how chromatin folding enables contact between gene promoters and targeted non-coding regulatory elements is not known. Genome wide mapping of chromatin interactions (HiC) and histone PTMs (ChIP-seq) demonstrated that while large scale higher order chromatin structures are stable across differentiation state, there were changes in genome contacts at a sub domain level at lineage specific gene loci. Interestingly, we identified that chromatin structures within killer T cells prior to infection (termed naive) exhibited a pre-configured structure, both at the level of histone PTMs and higher order chromatin contacts. This genomic pre-configuration was then associated with targeted epigenetic maturation of lineage-specific genomic elements upon T cell activation and acquisition of lineage-specific function, thus implying that the outcome of $CD8^+$ T cell differentiation is largely pre-determined. These data have implications for the molecular events, and their regulation, that occur during the generation of effective T cell responses and establishment of immunological memory.

1. 1. Russ et al., Regulation of H3K4me3 at transcriptional enhancers characterizes acquisition of virus-specific CD8+ T cell lineage-specific function. Cell Reports, in press.
2. 2. Russ et al., Mapping histone methylation dynamics during virus-specific CD8+ T cell differentiation in response to infection. Immunity. 41:853-865, 2014.

240

Highly Parallel Concurrent Gene Expression and Iso-Seq Analysis of Single Cells

Jafar S. Jabbari[1], Kirby Siemering[1], Camden Lo[2], Tu Nguyen[3, 4], Lucy Fang[3, 4], Alex W. Hewitt[3, 4, 5], Raymond Wong[3, 4]

1. Australian Genome Research Facilty, Parkville, VICTORIA, Australia

2. Private Consultancy, Melbourne

3. Opthamology, Department of Surgery, University of Melbourne , Melbourne, Victoria, Australia

4. Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, Melbourne, Victoria, Australia

5. School of Medicine, Menzies Institute for Medical Research, University of Tasmania, Hobart, Tasmania, Australia

Recent technological innovations have enabled genomic analysis of single cells and identifying role of genes and mechanisms regulating their expression in various biological processes such as cell differentiation. Among the techniques, droplet based single cell RNA-Seq has advanced rapidly allowing parallel analysis of tens of thousands of cells from a population using short reads from 3' or 5' end of transcripts. This approach has been very useful for clustering cells into various types based on gene expression signatures. However, in humans, majority of genes are spliced or transcribed alternatively giving rise to isoforms with different and sometimes opposing biological roles. Consequently, current high throughput methods lack the potential to investigate biological role or preferential expression of isoforms in different cell types.

To unveil the diversity of transcripts isoforms and their role in cell differentiation and other biological processes it is essential to cost effectively and concurrently characterise full-length transcripts of single cells along with quantitative expression analysis. To achieve this we have developed a method that takes advantage of 10x Genomics Chromium single cell solution to package cells in droplets for barcoding transcripts. We use cell-specific barcoded cDNA resulting from reverse transcription of polyA RNA content of a single-cell for standard 10x analysis using Illumina short read sequencing, and examine full-length transcripts of the same cells with PacBio Sequel platform. In this work we report application of the method to human post-mortem retina cells and discuss library characteristic and analysis results.

241

DECENT: Differential Expression with Capture Efficiency adjustmeNT for Single-Cell RNA-seq Data

Agus Salim[1, 2], Terence P Speed[1], Chengzhong Ye[1]

1. Bioinformatics Division, The Walter and Eliza Hall Institute of Medical Research, Melbourne, VIC, Australia

2. Mathematics and Statistics, La Trobe University, Melbourne, VIC, Australia

Due to the small amount of starting material and the low capture efficiency of the current technologies, scRNA-seq data contains a large number of zero counts due to the dropout phenomenon. The unique features of scRNA-seq data have led to the development of novel methods for differential expression (DE) analysis. However, few of the existing DE methods for scRNA-seq data estimate the number of molecules pre-dropout and therefore do not explicitly distinguish technical and biological zeroes. We develop DECENT, a DE method for scRNA-seq data that adjusts for the imperfect capture efficiency by estimating the number of molecules pre-dropout. Using simulated and real datasets, we saw the performance of our method is better compared to previously published methods, especially for detecting DE genes with low abundance. DECENT uses raw UMI-count data as input and does not require spike-ins, but when spike-ins are available, they can be used to improve its performance.

242

Comparative phylogenomic evidence for a novel detoxification gene family in insects

Jack Scanlan[1], Rebecca Smith[1], Philip Batterham[1, 2], Charles Robin[1]

1. School of BioSciences, The University of Melbourne, Melbourne, Victoria, Australia

2. Bio21 Molecular Science & Biotechnology Institute, Melbourne, Victoria, Australia

Xenobiotic metabolism in insects is thought to have many similarities to that of mammals but is comparatively poorly understood, despite its crucial relevance to plant-insect ecological interactions, agricultural insecticide use, and the suitability of insects as pharmacological models. Phosphorylation of xenobiotic compounds, a rare detoxification reaction in mammals, is widespread in insects, but the enzymes responsible are unknown. The EcKinases, a family of small-molecule kinases present in insects and crustaceans but not vertebrates, have been previously implicated in steroid hormone metabolism, but we here present comparative phylogenomic evidence suggesting some members are responsible for xenobiotic phosphorylation reactions and are therefore non-canonical detoxification enzymes. We have manually annotated the EcKinase family in 129 insect genomes and reconstructed their evolution within and between 11 insect orders. EcKinase family size varies dramatically across insect genomes, from 12 genes in most bees to 104 genes in the German cockroach, and positively correlates with the size of canonical detoxification gene families. Many clades of EcKinases in the saprophagous fly genus *Drosophila* have experienced rapid gene gain and loss, and their members tend to have enriched expression in digestive and metabolic tissues and are transcriptionally induced by xenobiotic compounds, strongly suggesting roles in detoxification. In contrast, bees and tsetse flies, two taxa with reduced need for xenobiotic detoxification, have comparatively fewer EcKinases, nearly all of which are very stable. In Lepidoptera, generalist herbivores have significantly more EcKinases than specialists, which can be explained by repeated, independent expansions of two ancestral clades in generalist lineages. Our data suggest EcKinases can contribute to xenobiotic metabolism in herbivorous and saprophagous insects, and we are carrying out genetic and toxicological experiments in *Drosophila melanogaster* to test their function in the detoxification of plant and fungal secondary metabolites.

243

Breaking point: computational interrogation of structural variation in cancer

Tracy Ballinger, Britta Bouwman, Nicola Crosetto, Colin A Semple

Structural variants (SVs) are known to play important roles in a wide variety of cancers, but their mutational origins and functional consequences are still poorly understood. The highly nonrandom distributions of these variants across tumour genomes are often assumed to reflect selective processes, but mutation rates can vary by orders of magnitude and often reflect the underlying chromatin structure at a locus. The inference of SVs under selection for enhanced tumourigenesis therefore remains challenging, though identifying such variants may lead to new diagnostic and therapeutic targets. Using experimentally derived mutation data we derive the first quantitative models of double strand break (DSB) frequency across the human genome, based upon underlying chromatin and sequence features. These models provide high predictive accuracy, and models trained in one cell type can be successfully applied to others. We show that most SV 'hotspots' (harbouring unusually high SV breakpoint frequencies) seen across a variety of tumour sequencing studies are broadly consistent with DSB model predictions. Using model predictions as a proxy for susceptibility to mutation in tumours, many SV hotspots appear to be adequately explained by selectively neutral mutational bias alone. However, a fraction of hotspots show SV breakpoint frequencies that are unexpectedly high given their predicted susceptibility to mutation, and are therefore credible targets of positive selection in tumours. In contrast, hundreds of regions across the genome show unexpectedly low levels of SVs, given their relatively high susceptibility to mutation. These novel 'coldspot' regions appear to be subject to purifying selection in tumours. Both the hotspot and coldspot regions predicted in this manner show intriguing enrichments for genes and regulatory elements.

244

Investigating the role of R-loops in erythroid differentiation

Manan Shah[1], Kate GR Quinlan[1], Merlin Crossley[1]

1. School of Biotechnology and Biomolecular Sciences, UNSW Sydney, Kensignton, NSW, Australia

Transcriptional regulation is crucial for a number of biological processes and involves the binding of sequence-specific transcription factors (TFs) which in turn affect the recruitment of RNA polymerase II and the rate of gene expression. This process is now known to involve many other layers of regulation such as the structure of chromatin. Genomic DNA can adopt many different structures apart from the usual B-form double stranded DNA (dsDNA), one such structure being an R-loop.

R-loops are formed when dsDNA is separated and a complementary ssRNA is able to bind one of the DNA strands, resulting in a RNA/DNA hybrid and a region of ssDNA. The development of recent techniques such as DNA-RNA immunoprecipitation (DRIP), a technique analogous to ChIP have shown that these structures are abundant throughout the mammalian genome. Furthermore, specific R-loops have bene shown to act as mediators of transcriptional regulation and cellular differentiation. These structures are particularly enriched in promoter regions, leading us to hypothesise that they play a role in transcriptional regulation and, more specifically, in modulation of TF binding.

To investigate this potential role, we will first map the location of R-loops genome wide in erythroid cell lines and combine this with publicly available data to identify regions of interest. We will then aim to target these structures to identify effects that R-loops have on TF binding, chromatin accessibility and/or gene expression.

245

Developing a bioinformatics pipeline to measure changes in ribosomal RNA genes copy number in cancer

Diksha Sharma[1], Sylvie Hermann Le Denmat [2, 3, 1], Justin O'Sullivan[4], Katherine M. Hannan[5], Ross D. Hannan[5], Austen R.D. Ganley[6, 1]

1. School of Biological Sciences, University of Auckland, Auckland, New Zealand

2. Ecole Normale Superiere-Paris, Paris, France

3. Institute for Integrative Biology of the Cell, Paris, France

4. Liggins Institute, University of Auckland, Auckland, New Zealand

5. Department of Cancer Biology and Therapeutics, Australian National University, Canberra, Australia

6. Maurice Wilkins Centre for Molecular Biodiscovery, University of Auckland, Auckland, New Zealand

Cancer is a leading cause of death worldwide. A hallmark characteristic of cancer is abnormal nucleolar morphology. Nucleoli house the ribosomal RNA genes (rDNA) that encode ribosomal RNA, which is the major structural and catalytic component of ribosomes. In eukaryotes, the rDNA is organized as tandem arrays of repeats that exhibit a high degree of variability in copy number within and between species.

Recent evidence suggests that rDNA copy number changes in malignancy. However, determining rDNA copy number in mammals is technically challenging, and most methods have not been properly validated. The current dominant method uses whole genome sequencing read coverage (or depth) as a proxy for rDNA copy number. This method assumes that average coverage represents the true coverage value across both the rDNA and the whole genome. However, there are regions with high or low coverage across the rDNA repeat unit, presumably due to the presence of interspersed repeats (Alu elements), tandem repeats (microsatellites), and sequencing bias. This coverage variability might result in inaccurate estimation of the final average coverage level, hence giving a false estimation of the rDNA copy number.

To overcome these limitations, we have developed an approach that uses the most frequent coverage value to calculate copy number. The methodology is based on the assumption that the most frequent coverage value represents the true coverage value. We have validated our method using yeast strains with varying, known rDNA copy numbers. This validated system will be employed using paired cancer-normal whole genome sequence data to measure the variation in rDNA copy number between different individuals and between different tissue types. This will enable us to establish the normal human variation in rDNA copy number, and to assess whether the rDNA copy number changes previously observed in malignant cells fall outside this normal range of variation.

246

In vivo targets and mechanisms of gene control during protein synthesis

Nikolay Shirokikh[1], Stuart Archer[2], Susan Wagner[1], Yoshika Janapala[1], David Powell[2], Kate Hannan[3], Rick Pearson[4], Ross Hannan[3, 4], Traude Beilharz[2, 5], Thomas Preiss[1, 6]

1. Department of Genome Sciences, The John Curtin School of Medical Research, The Australian National University, Canberra, ACT, Australia

2. Monash Bioinformatics Platform, Monash University, Melbourne, Vic, Australia

3. Department of Cancer Biology & Therapeutics, The John Curtin School of Medical Research, The Australian National University, Canberra, ACT, Australia

4. Division of Cancer Research, Peter MacCallum Cancer Centre, Melbourne, Vic, Australia

5. Monash Biomedicine Discovery Institute and Department of Biochemistry and Molecular Biology, Monash University, Melbourne, Vic, Australia

6. Victor Chang Cardiac Research Institute, Sydney, NSW, Australia

Regulation of protein synthesis on mRNA is central to eukaryotic expression of genes. Regulatory inputs are specified by the mRNA untranslated regions (UTRs) and often target translation initiation. Initiation consists of dynamic and complex interactions between eukaryotic initiation factors (eIFs) and the small ribosomal subunit (SSU), and concludes with joining of the large ribosomal subunit (LSU) to form a full ribosome and translate the mRNA code. While control of protein synthesis is important for fast-paced cell adaptation, methods to study the mechanistic intricacies of translation initiation *in vivo* transcriptome-wide were lacking.

We developed translation complex profile sequencing (TCP-seq) [Archer SK et al. Nature 2016 535:570-574; Shirokikh NE et al. Nat. Protoc. 2017 12:697-731], a method related to the ribosome profiling approach. TCP-seq uniquely allows to resolve all translation intermediates, including the elusive mRNA 'scanning' by SSUs, visualize start codon recognition events, and capture diverse conformations of elongating ribosomes *in vivo.* Thus, the method can be used to pinpoint mRNAs with high potential of functional control and directly visualize locations of the regulatory elements in their 5'UTRs.

Combining TCP-seq with eIF-selective purification of complexes in yeast model system, we aim to discern the fundamental mechanistic problems of initiation, such as what proteins confer directionality to the SSUs during scanning, and how ribosomal recruitment to mRNA is controlled during nutrient stress conditions. We further adapt TCP-seq to use in mammalian cells and aim to obtain snapshots of translation in normal and cancer cells to identify 5'UTRs which can be selectively up- or downregulated in malignancies.

247

Targeting telomeric repeat-binding factor 2 (TRF2) with small molecule inhibitors to prevent cancer cell growth

Alexander P. Sobinoff[1], Ronnie R. J. Low[2], David Van Ly[2], Keriann M. Backus[3], Anthony J. Cesare[2], Sandro Cosconati[4], Hilda A. Pickett[1]

1. Telomere Length Regulation Unit, Children's Medical Research Institute, University of Sydney, Sydney, NSW, Australia

2. Genome Integrity Unit, Children's Medical Research Institute, University of Sydney, Sydney, NSW, Australia

3. The Skaggs Institute for Chemical Biology and Department of Molecular Medicine, The Scripps Research Institute, La Jolla, CA, USA

4. DiSTABiF, University of Campania Luigi Vanvitelli, Naples, Italy

Telomeres are specialised DNA-protein structures which cap the ends of chromosomes. Telomeres are bound by the protein complex shelterin, which prevents the aberrant activation of a DNA damage response at chromosome ends. The telomere binding protein TRF2 is a constituent member of the shelterin complex, and is the primary inhibitor of ATM DNA damage signalling at the telomere. Cancer cells have several telomere deficiencies brought about by replication stress, resulting in fragile telomeres and an elevated DDR at the telomere. These deficiencies represent a unique vulnerability which can be exploited for the treatment of cancer. Targeting TRF2, the guardian of the telomere, with small molecule inhibitors which block its ability to protect the telomere represents a unique avenue for the development of targeted cancer therapy. We have synthesised several cell permeable small molecule inhibitors which we have shown directly interact with the TRFH domain of TRF2 *in vivo*, causing a telomere-specific DNA damage response in cancer cells without affecting mortal controls. This increased DNA damage response coincided with a reduced cell viability specifically in cancer cells. Therefore, cancer cells are potentially more sensitive to the chemical inhibition of TRF2 domains compared to healthy cells, making TRF2 a potentially powerful therapeutic target to kill cancer cells.

248

Gene editing of the multi-copy H2A.B.3 gene family by a single pair of TALENs

Tanya Soboleva[1], Diana Anuar[1], Josephine Bowles[2], Peter Koopman[2], Philip Gregory[3], Matt Field[4], Sebastian Kurscheid[1], David Tremethick[1]

1. JCSMR, ANU, Canberra, ACT, Australia

2. School of Biomedical Sciences , Faculty of Medicine, UQ, Brisbane, Queensland, Australia

3. Sangamo BioSciences Inc, Richmond, , California, , USA

4. Centre for Biodiscovery and Molecular Development of Therapeutics, James Cook University, Townsville, Queensland, Australia

In view of the controversy related to the generation of off-target mutations by gene editing approaches, we tested the specificity of TALENs by disrupting a multi-copy gene family of histone variant, H2A.B3, using only one pair of TALENS. This represents the first described knockout of this histone variant.

We are showing that all three mutations are stably inheritable up to the present, 9th generation of KO mice. Most importantly, we have investigated, using exome sequencing, whether TALEN activity induced any off-target mutations in our mouse model. For that, we sequenced the genomes of mice that belong to 3 consecutive generations and compared the results to wild type mice that were used for breeding of our colony. Our results showed that TALEN technology is capable of producing extremely well-targeted mutations in multi-copy genes and did not produce off-target effects that could be identified by exome sequencing. We also showed that wet-lab validation of predicted off-target sites is crucial in order to make valid conclusions about the specificity of targeting. The phenotypic characterization of a H2A.B3 KO and the impact on the fertility and the epigenome of male germ cells that lack H2A.B.3 will also be presented.

249

Investigating the role of *MYD88* mutations in lymphoma.

Huma Sohail, James Wang, Naomi Mitchell, Olga Zaytseva, Keisuke Horikawa, Leonie Quinn

Lymphoma is the sixth most common form of cancer, with 90% of cases belonging to a heterogeneous group of lymphomas called Non-Hodgkin Lymphoma (NHL). One of the most common types of NHL, Diffuse Large B Cell Lymphoma (DLBCL), is characterized into two subtypes on the basis of differential gene expression: Activated B Cell (ABC) and Germinal Centre (GC), of these ABC is the most aggressive and least curable form. Recurrent somatic mutations in *MYD88* were initially identified in ABC-DLBCL patients. MYD88 is an adaptor protein that transmits innate immune signalling from Toll-like receptor (TLR) and interleukin-1 (IL-1) receptor to the nuclear factor-kappa B (NF-kB) pathway, for immune and inflammatory responses. The most frequent *MYD88* mutation in lymphoma is a single amino acid variant L265P, identified in approximately 30% of ABC-DLBCL and 90% of Waldenström Macroglobulinemia (WM) patients. Apart from the L265P mutation, additional mutations in *MYD88* have been reported in lymphoma. Moreover, genome-wide analysis of ABC-DLBCL and WM indicates that MYD88 L265P is preferentially found in combination with mutations in the B-cell antigen component CD79B and chemokine receptor CXCR4, which result in further induction of NF-κB pathway.

We are currently investigating the role of *MYD88* in both normal development and immune cell function. For this purpose, we have developed *Drosophila melanogaster* and mouse model systems. In the fly model, we are investigating potential effects of *Myd88* knockdown and CRISPR-generated Myd88 I335P substitutions (designed to mimic oncogenic MYD88 L265P) on hematopoietic development and immunity. In the mouse model, compound mutations (MYD88 loss-of-function together with MYD88 L265P) are being examined to determine potential of MYD88 L265P as a therapeutic target. Further to this, we are testing the combined effect of the MYD88 L265P mutation and CD79B mutations in our B cell lymphoma mouse models.

250

BET bromodomain inhibition sensitises ribosomal DNA to localised damage and provides a robust synergistic therapeutic strategy for treating AML

Jirawas Sornkom[1, 2, 3], Donald Cameron[1, 2, 3], Nadine Hein[1, 3], Natalie Brajanovski[1], Chun Yew Fong[1], Omer Gilan[1], Elaine Sanij[1], Kate Hannan[3], Neil Garton[4], Rabinder Prinjha[4], Mark Dawson[1, 2], Carleen Cullinane[1], Rick Pearson[1, 2, 5, 6], Ross Hannan[1, 2, 5, 7, 3, 6], Gretchen Poortinga[1, 8, 2]

1. Research Division, Peter MacCallum Cancer Centre, Melbourne, VIC, Australia

2. Sir Peter MacCallum Department of Oncology, University of Melbourne, Parkville, Victoria, Australia

3. ACRF Department of Cancer Biology and Therapeutics, John Curtin School of Medical Research, Canberra, ACT, Australia

4. Medicines Research Centre , GlaxoSmithKline R&D, Hertfordshire, UK

5. Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria, Australia

6. Department of Biochemistry and Molecular Biology, University of Melbourne, Parkville, Victoria, Australia

7. School of Biomedical Sciences, University of Queensland, Brisbane, Queensland, Australia

8. Medicine, St Vincent's Hospital, University of Melbourne, Fitzroy, Victoria, Australia

RNA polymerase I (Pol I) mediated transcription of the ribosomal RNA genes (rDNA) is frequently upregulated in cancer, including acute myeloid leukemia (AML)[1,2]. rDNA is highly vulnerable to DNA damage and inhibition of Pol I transcription by a selective inhibitor CX-5461, which is currently in clinical trials, induces a nucleolar-specific DNA damage response (DDR) via ATM/ATR signalling. While single agent CX-5461 provides an impressive survival benefit in preclinical mouse models of AML, animals eventually succumb to disease, underscoring the need for Pol I inhibitor combination therapies[1,2]. Inhibitors of bromodomain and extra-terminal (BET) proteins, also in clinical trials, block BET protein binding to acetylated chromatin marks and growing evidence has established a role for BET proteins in DDR regulation, via transcriptional regulation and direct effects on chromatin accessibility[3,4]. Here, we report a striking synergy between CX-5461 and the BET protein inhibitor IBET-151, where their co-treatment significantly improved survival over single agents in preclinical models of aggressive AML. Following treatment of AML cells with IBET-151 or the knockdown of its target BRD4, we demonstrate a dramatic increase in rDNA accessibility to MNase. Significantly, co-treatment of AML cells with IBET-151 and CX-5461 led to a heightened DDR at rDNA chromatin, evidenced by marked γH2AX enrichment at rDNA. Mechanistically, our data demonstrate that inhibition of BRD4 by IBET-151 results in more open rDNA chromatin, exposing the rDNA to enhanced CX-5461-mediated DDR and leading to synthetic lethality of AML cells. Intriguingly, IBET-151-resistant cells[5] are re-sensitised to BRD4 inhibition when combined with CX-5461, where co-treatment activated this synergistic ATM-mediated response independent of IBET-151 resistance pathways. Together, these studies highlight the therapeutic value of using BET inhibitors to amplify the underlying sensitivity of rDNA to DNA damage and thus exploit a previously unexplored synthetic lethality between BET protein and Pol I transcription inhibition for treating AML.

1. Bywater et al., Cancer Cell, 2012
2. Hein et al., Blood, 2017
3. Floyd et al., Nature, 2014
4. Yang et al., Sci Trans Med, 2017
5. Fong et al., Nature, 2015

251

Nuclear microRNAs as Direct Regulators of Transcription

Laura Sourdin[1], Katherine Pillman, John Toubia, Klay Saunders, Greg Goodall, Cameron Bracken

1. University of South Australia; Centre for Cancer Biology, SA, Adelaide BC, SA, Australia

MicroRNAs are well established suppressors of gene expression at the post-transcriptional level, targeting genes through complementary base pairing then repressing gene expression at the levels of both transcript stability and translation. Recent data from our lab and others however indicate the main response of cells to microRNAs are in-fact transcriptional. We hypothesise many of these effects are explained by the direct regulation of transcription factors by microRNAs, and the knock-on effects which then occur as a result of this primary interaction. Although many studies dismiss such responses as “indirect”, these effects are in-fact central to microRNA function, with microRNA : transcription factor co-regulatory loops being an over-represented motif within genetic networks. I present both bioinformatic and preliminary experimental data identifying such interactions that are likely important components of the miR-200 regulatory network – a critical microRNA that controls Epithelial-Mesenchymal Transition during both development and metastasis. It is likely these responses are not unique to miR-200, but rather reflect a more generalised property of microRNA function.

252

Deep-Mpute: Imputation of missing methylation values using deep convolution neural networks

Akanksha Srivastava[1], Akshay Asthana[2], Justin Borevitz[3], Ryan Lister[1, 4]

1. ARC Centre of Excellence in Plant Energy Biology, University of Western Australia (UWA), Perth , WA, Australia

2. Seeing Machines , Canberra, ACT, Australia

3. Research School of Biology (RSB), Australian National University, Canberra, ACT, Australia

4. Harry Perkins Institute of Medical Research, 6 Verdun St, Nedlands, WA, Australia

The development of whole genome sequencing techniques has made it possible to quantify DNA methylation levels at single-base resolution across an entire genome. However, due to the high cost of sequencing, missing or insufficient coverage at individual CpG sites remains a major issue. To tackle this critical problem, we have developed a user-friendly tool for the imputation of methylation values at single CpG sites using a deep convolution neural network (CNN). We demonstrate that our tool overcomes the biases present in current DNA methylation data resulting from variability in sequencing depth across samples and hence can be successfully used for imputing low or missing methylation values before conducting any further analysis, such as the identification of differentially methylated regions. The CNN itself is trained by integrating important features such as neighboring CpG methylation values, genomic distance and the methylation levels from available replicate data. We show that Deep-Mpute can reliably predict the missing methylation values in both plant and mammalian genomes and outperforms current state-of-the-art tools.

253

The Search for Disease Severity Modifiers in Patients Suffering from Blood Disorders

Elizabeth S Stout[1], Kate GR Quinlan[1], Merlin Crossley[1]

1. School of Biotechnology and Biomolecular Sciences, UNSW Sydney, KENSINGTON, NSW, Australia

Blood disorders such as β-thalassemia and sickle-cell anaemia arise due to mutations in the genes encoding the β-globin subunits of haemoglobin. Haemoglobin is a tetrameric heterodimer composed of two α-like and β-like subunits present in red blood cells that is essential for carrying oxygen around the body. To cater for varying oxygen demands during development, the composition of haemoglobin changes in a tightly-regulated process known as globin switching. This involves two switches in the composition of β-globin, first from embryonic-to-foetal globin during gestation, and secondly from foetal-to-adult after birth. Due to compelling evidence suggesting that increased foetal-globin can compensate for loss of functional adult-globin, studies have aimed to uncover the mechanisms behind globin switching as a means to 'reawaken' the foetal genes therapeutically.

GWAS studies have identified three major regions that are associated with high foetal-globin levels, however, these sites cannot account for the variation in foetal-globin expression between humans. In order to expand our knowledge on the variants involved with foetal-globin regulation, we aim to identify additional genetic variants that are involved with the regulation of globin genes. Regions of interest include those within the *β-globin* locus, such as the 5'HS4-LCR polymorphism, as well as in the genes encoding factors known to be important in globin regulation, such as *Klf1*. Our approach is to engineer variants of interest into erythroid progenitor cell lines by CRISPR/Cas9 in order to characterise the impact of these variants on foetal-globin expression. Future work will involve analysis of globin expression at the mRNA level by RT-PCR and at the protein level by flow cytometry.

We aim to utilise our growing understanding of the importance of these genetic variants to aid in informing novel therapeutic approaches to the treatment of blood disorders such as β-thalassemia and sickle-cell anaemia.

254

Invasive Lobular Breast Cancer: Using tumour genome-wide DNA methylation to further subtype and aid in the identification of susceptibility genes.

Medha Suman[1], JiHoon Eric Joo[2, 1], Ee Ming Wong[2, 1], Tu Nguyen-Dumont[2, 1], Neil O'Callaghan[1], Melissa Yow[1], John Hopper[3], Graham Giles[3, 4, 5], Roger Milne[3, 4], Melissa C. Southey[2, 1, 4]

1. Department of Pathology, Genetic Epidemiology Laboratory, Victorian Comprehensive Cancer Centre, The University of Melbourne, Melbourne, Victoria, Australia

2. Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Melbourne, Victoria, Australia

3. Centre for Biostatistics and Epidemiology, School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia

4. Cancer Epidemiology and Intelligence Division, Cancer Council Victoria, Melbourne, Victoria, Australia

5. School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia

There is indirect evidence of an inherited component to risk of invasive lobular breast cancer (LBC). We and others have conducted a whole-genome sequencing project involving 120 population-based early-onset and clinic-based multiple-case LBC cases aimed at identifying susceptibility genes for LBC. The project has

identified a vast amount of germline genetic variation but no strong candidate LBC susceptibility gene. The role of epigenetic alterations in the susceptibility and progression of cancers has been widely accepted. Here we have sought to identify methods to subtype LBC into groups of reduced heterogeneity using genome-wide DNA methylation patterns in order to interpret the whole genome sequences more precisely.

Formalin-fixed paraffin-embedded (FFPE) LBC-enriched DNA was prepared from 161 LBC samples using macrodissection and run on the Infinium HumanMethylation450K Beadchip (HM450K) array to generate genome-wide DNA methylation data. The raw methylation data was pre-processed and normalised using *minfi* Bioconductor package in R programming software. Unsupervised cluster analysis was used to identify subtypes of LBC based on DNA methylation involving 449005 probes. This dataset was also compared with similar data prepared from FFPE samples of all other breast cancer subtypes (n=341).

LBC samples were found to cluster into three main subtypes. We will present data that describes the analyses that have identified differentially methylated genomic regions between these three LBC subtypes. The identification of LBC subtypes using DNA methylation that increase the homogeneity of disease in each subtype could inform the analysis of the whole-genome sequencing project and lead to the identification of LBC susceptibility genes.

255

Integrated profiling of single-cell chromatin accessibility and transcriptome reveals regulatory heterogeneity

Longqi Liu[1, 2, 3], Chuanyu Liu[1, 2, 4], Liang Wu[1, 2], Yue Yuan[1, 2, 4], Mingyue Wang[1, 2, 4], Fang Chen[1, 2], Zhouchun Shang[1, 2, 5], Xun Xu[1, 2], Stephanie Sun[1]

1. BGI-Shenzhen, Shenzhen, China

2. China National GeneBank, BGI-Shenzhen, Shenzhen, China

3. Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, China

4. BGI Education Center, University of Chinese Academy of Sciences, Shenzhen, China

5. Department of Regenerative Medicine, Tongji University School of Medicine, Shanghai, China

The advances of single cell sequencing technologies have greatly improved our understanding of heterogeneity in terms of genetic, epigenetic and transcriptional regulation within cell populations. We and other groups have developed single-cell whole genome, exome, methylome and transcriptome technologies and applied these approaches to analyzing the complexity of cell populations in tumorigenesis, developmental process and cellular reprogramming. Meanwhile, single-cell epigenome techniques including single cell ChIP-seq, ATAC-seq, DNase-seq and Hi-C, have been developed to decipher histone modifications, chromatin accessibility landscapes, and 3D chromatin contacts respectively in single cells. Integrative analysis of single-cell multimodal data is critical for accurate dissection of cell-to-cell variation within certain cell populations. Recent progress on measuring multi-omics in the same cells has enabled analysis of associations between different layers of regulation on gene expression. So far, the relationship between chromatin accessibility and gene expression has not been investigated at the resolution of one single cell. To address this question, we report scCAT-seq, a technique for simultaneous assay of chromatin accessibility and transcriptome within the same single cell. By applying scCAT-seq to different cancer cell types, we identified *trans*-factors as bridges linking accessibility variation of *cis*-regulatory elements to cell-type-specific gene expression across single cells. We further characterized subpopulations within cancer cells and uncovered the regulatory clues that drive transcriptional heterogeneity. Together, scCAT-seq is a promising tool for the joint analysis of multimodal data of single cells, which also offers the potential for clinical applications such as preimplantation screening and cancer diagnosis.

256

Exploring the combinatorial effects of epigenetic modifiers upon the mammalian genome using CRISPR-dCas9

Tessa Swain[1, 2], Ryan Lister[1, 2], Christian Pflueger[1, 2], Trung Nguyen[1, 2]

1. The University of Western Australia, Crawley, WA, Australia

2. Epigenetics and Genomics, The Harry Perkins Institute of Medical Research, Nedlands, WA, Australia

The epigenome is comprised of chemical modifications to DNA and histone proteins that regulate chromatin structure and DNA accessibility. Given the variety and complexity of epigenetic modifications, and the diverse repertoire of enzymes that regulate them, understanding the role of these modifications in transcriptional regulation will require comprehensive exploration of their combinatorial activities. While first generation

epigenome engineering tools using single catalytic domains allow targeted editing of DNA and histone modifications, they frequently underperform in efficacy and stability. Consequently, we will deeply probe the combinatorial effects of recruiting multiple epigenome modifying enzymes simultaneously to target loci in the human genome, investigating their causal roles in genome regulation. This will provide new insights into how epigenome modification pathways function to regulate gene expression, while developing a range of novel epigenome editing systems for synthetic biology applications.

257

Identifying Novel Testis-Determining Genes from Integrated RNAseq and ChIPseq Data

Aleisha Symon[1, 2], Rowena Lavery[1], Massilva Rahmoun[3], Janelle Ryan[1], Dimuthu Alankarage[1], Stefan Bagheri-Fam[1], Francis Poulat[3], Vincent Harley[1, 2]

1. Hudson Institute of Medical Research, Clayton, VIC, Australia

2. Department of Anatomy and Developmental Biology, Monash University, Melbourne, Victoria, Australia

3. Institute of Human Genetics, University of Montpellier, Montpellier, France

SOX9 is a key transcription factor and Sertoli cell fate determinant responsible for the differentiation of the gonad into a testis during embryonic development. Human *SOX9* mutations cause Disorders of Sex Development (DSD) in XX males (*SOX9* duplications) and XY females (*SOX9* mutations/deletions), however most DSD patients don't receive a definitive genetic diagnosis, making clinical management challenging. We hypothesise that SOX9 target genes are candidate DSD genes.

To identify Sox9 target genes, we undertook RNAseq analysis on mouse *Sox9* knock-out gonads from embryonic day E13.5, when *Sox9* is ablated in an intact Sertoli cell environment. We also performed Sox9 ChIPseq on wildtype E13.5 mouse testes and E90 bovine testes. 240 genes were downregulated in the *Sox9* knockout testes, thus activated by Sox9. 4293 Sox9 ChIPseq peaks were common to the mouse and bovine testis. Overlapping the RNAseq and conserved ChIPseq datasets identified 119 genes whose gonadal chromatin is bound by Sox9, and whose gene expression is upregulated by Sox9. Of these, 34 have an enriched expression in Sertoli Cells compared to the other cell types of the gonads. We are now elucidating the role of these genes in the development of the testis.

One of these genes is *Nedd9*, a scaffolding protein in the Cas family which is involved in many cellular processes including cell polarity and migration. By manipulating *SOX9* levels in a human Sertoli Cell-like culture model, we have also shown that this regulation is likely relevant to humans too. Initial analyses of *Nedd9* knockout embryos has revealed abnormal testis morphology, including a mis-located ceolomic blood vessel, a male-specific feature of gonadal vasculture.

These studies may reveal a previously unsuspected role of Sox9 in the development of the vasculature of the testis and may provide another candidate gene for screening of DSD patients.

1. Rahmoun, Massilva, et al. "In mammalian foetal testes, SOX9 regulates expression of its target genes by binding to genomic regions with conserved signatures." Nucleic Acids Research (2017).

258

Investigating the role of Cis and Trans acting regulators in the process of X inactivation

Andres ATdF Tapia del Fierro[1], Natasha NJ Jansz[1], Megan MI Iminitoff[1], Kelsey KB Breslin[1], Jamie LJG Gearing, Tracy TW Wilson[1], Matthew MER Ritchie[1], Andrew AK Keniry[1], Marnie MEB Blewitt[1]

1. Walter and Eliza Hall Institute of Medical Research, West Melbourne, VIC, Australia

X chromosome inactivation (XCI) is the most important silencing event undergone by female mammalian cells. Virtually all known epigenetic silencing mechanisms come into play during this process, which makes it a great model system to study them. Among the actors involved, some are *trans*-acting, such as the polycomb repressive complexes (PRC) 1 and 2, and some are *cis*-acting like the long non-coding RNA Xist that coats the inactive X chromosome. To find new *trans* epigenetic modifiers, we perform high-throughput screens in female Xmas (X-linked marker active silent) embryonic stem cells undergoing X inactivation. These cells possess a different reporter gene (GFP or mCherry) in the Hprt allele of the X chromosome that becomes silenced with the rest of the chromosome, thus allowing the monitoring of XCI by FACS. Target genes are then knocked-down by shRNAs to see if this interferes with X gene silencing.

On the other hand, *cis*-acting effectors include structural elements like the Dxz4 macrosatellite repeat, which marks the boundary between the two inactive X's mega domains. Other repetitive elements such as tandem or interspersed repeats also seem to have roles in the three-dimensional conformation of the inactive X. One of such is the locus of the repetitive long non-coding RNA Firre. Although the transcript RNA has been reported to have different roles, the genomic locus has a seems to be involved in setting the inactive X's characteristic

conformation. This locus has been showed to co-localize with Dxz4 only in the inactive X and the nuclear periphery.

During my PhD, I plan on using screening techniques to discover new *trans*-acting epigenetic modifiers, and to study *cis*-acting repetitive elements by using an array of different genomic techniques.

259

Investigating DNA transposition as a cause of genomic instability and therapy-resistance in Chronic Myeloid Leukaemia

Daniel Thomson[1], Paul Wang[1], Andreas Schreiber[1], Susan Branford[1]

1. Centre for Cancer Biology, Adelaide, SA, Australia

Genomic instability is a hallmark of cancer, providing a means by which cancer cells can adapt to selective pressures and acquire drug-resistance. In the treatment of chronic myeloid leukaemia (CML), a subset of patients develop resistance to treatment with tyrosine kinase inhibitors (TKI's). We investigate DNA transposition as a source of genomic instability in CML, identifying the RAG (recombination activation genes) as potential drivers of transformation of CML. RAG1 has recently been shown to have evolved from DNA transposons but are well known for their essential role in acquired immunity; RAG1 is physiologically expressed in developing lymphocytes where it cuts and pastes immunoglobulin genes together to enable antibody diversity. The potential for RAG1 to act as a transposon has been demonstrated *in vitro,* but its ability to contribute widely to genomic instability is only recently being recognised with application of high throughput sequencing strategies.

We have applied, RNAseq and whole exome sequencing (WES) on a cohort (n=186) of CML patient samples, in an effort to identify drivers of therapy-resistance. Analysis of gene expression has revealed aberrant expression of RAG1 and associated genes, specifically in patients at lymphoid blast crisis (advanced CML). Through analysis of copy number variation (CNV), we observe loss of immunoglobulin loci in CML patients matching RAG expression, indicative of RAG activity. This is supported by breakpoint analysis showing DNA breaks occurring at known RAG target sites. We also identify hundreds of putative fusion genes that occur since diagnosis, from which unbiased motif enrichment searches reveal an enrichment for the known RAG1 target motif, and in many cases include immunoglobulin genes mobilised throughout the genome.

This work provides evidence that RAG1 mediated DNA transposition may be a major driving force behind genomic instability in advanced CML, and potentially the source of drug-resistance.

260

Bayesian neural network based modelling of steady-state splicing mechanism.

KANUPRIYA TIWARI, Lars K Nielsen

Rationale: Most splicing models assemble a sequence based “splicing code” in a single tissue or rank sequence features based on their importance in different tissues making implicit assumptions about splice factor availability. From a systems biology perspective, the mechanistic modelling of splicing is complicated due to the expanding repertoire of core splicing machinery components and auxiliary factors involved. However, with the wealth of open source RNA-seq data available, machine learning methods can be trained using this data for learning the underlying mechanisms. We propose that by incorporating splice factor expression signatures and sequence features as inputs, a neural network model can be trained for predicting the most likely splicing pattern of a sequence, taking advantage of the methods ability to learn non-linear dependencies among its inputs.

Methods: We have built a 3 layer Bayesian neural network for predicting the direction of inclusion of the central exon in a triplet of exons. Each training data point for the network is a set of input features including 1400 sequence-based features and splice factor expression values (a linear combination of factor expression and its binding site counts) corresponding to 1 of 16 human tissues and the matched output probability of inclusion of the central exon in that tissue. A sparsity prior was assigned to the network weights for the purpose of regularizing the model, inference is performed using Gibbs sampling.

Impact: While models exist for predicting the effects of sequence variants on splicing, these variants only account for ca. 15% of disease-causing mutations whereas several diseases are a result of aberrant splice factor expression. The current model is a useful tool for assessing the impact of changes in factor expression on the transcriptome of a cell which can be used to narrow down the number of possibly affected transcripts to be further investigated.

261

REPETITIVE RNA AND GENOMIC INSTABILITY IN HIGH-GRADE SEROUS OVARIAN CANCER PROGRESSION AND DEVELOPMENT

James R Torpy[1], Nenad Bartonicek[1], Marcel E Dinger[3, 2, 1], David DL Bowtell[4, 1]

1. Garvan Institute of Medical Research, Darlinghurst, NSW, Australia

2. Kinghorn Centre for Clinical Genomics, Sydney, NSW, Australia

3. GenomeOne, Darlinghurst, NSW, Australia

4. Cancer Genomics and Genetics, Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia

Ovarian cancer is a highly complex disease with a range of different histological subtypes. This highly lethal disease is estimated to be the fifth most common cause of death from cancer in females, with a five-year relative survival rate of 46.2%. High-grade serous ovarian cancer (HGSOC), characterized by widespread genomic instability, accounts for 70-80% of ovarian cancer deaths, and survival rates have not improved significantly for the last few decades. Furthermore, the underlying cause of around 1/3 of HGSOC cases cannot be explained.

Evidence suggests that RNA derived from repetitive regions of the genome plays a role in genomic instability and development of cancers such as high-grade serous ovarian cancer, and may play a role in the unexplained HGSOC cases. Aberrant expression of centromere-derived RNA causes dysfunctional chromosomal segregation during mitosis and aneuploidy. Telomere-derived RNA maintains telomeres, preventing chromosomal fusion, breakage and subsequent rearrangement of the chromosomes. Retrotransposable elements such as LINE1s and *Alus* insert into different genomic locations, disrupting sequences and causing rearrangements such as duplications, inversions and translocations.

We have analysed over 120 HGSOC case and control RNA-sequencing data sets of primary samples from the Australian Ovarian Cancer Study, comparing differences in expression of repetitive RNA transcripts across multiple HGSOC subtypes and controls. We found a range of differentially expressed repetitive RNA species including LINE1, Alu and centromere-derived RNA which may be contributing to genomic instability in these tumours. In order to investigate the potential causes of the differences in repeat RNA levels, their expression was correlated with expression of a range of methyltransferases such as DNMT1 and DNMT3A-C that are known to regulate methylation at repetitive heterochromatin, controlling RNA expression from these regions. Expression of RNAi-associated factors such as Dicer was also assessed as these factors can contribute to repetitive RNA regulation.

262

Investigation into the mechanisms of alternative 3'UTRs in the control of gene expression

Rachael E Turner[1], Traude H Beilharz[1], Paul Harrison[2], Bernhard Dichtl[3], Angavai Swaminathan[1], Belinda Goldie[1]

1. Department of Biochemistry and Molecular Biology, Monash University, Melbourne, VIC, Australia

2. Monash Bioinformatics Platform, Monash University, Melbourne, VIC, Australia

3. Centre for Cellular and Molecular Biology, School of Life and Environmental Sciences, Deakin University, Burwood, Victoria, Australia

mRNA represents an important intermediate in protein synthesis. Cleavage and polyadenylation at the 3'end is carried out by complex machinery that is positioned through interactions with the mRNA transcript. When more than one potential poly(A) exists for a single gene this provides the opportunity for multiple mRNA isoforms to occur that may differ in their coding region or simply the length of their 3'UTR. This is referred to as alternative polyadenylation (APA) and has been linked to embryonic development, cell proliferation and multiple human diseases including cancer. The method by which alternative polyadenylation sites are chosen remains a relative mystery.

Temperature sensitive knockout of 3'end processing complex subunits generally induced a switch to longer mRNA transcripts in *Saccharomyces cerevisiae*. A similar lengthening was seen after treating cells with the adenosine anologue cordycepin. Furthermore genes that are able to undergo alternative polyadenylation tend to have a wider nucleosome-free region at their 3'end indicating that the ability to undergo APA is determined in part by the intrinsic higher ordering of the gene itself.

263

Pre-miRNA Folding Through Context-Free Grammar Parsing and the Identification of miRNA Using a Feedforward Neural Network

Viktor Prypoten[1], Sean Gribben[1], Xavier Pellow[1], Reena Zelenkova[1], David Helmerson[1], Joshua Nibbs[1], Boris Deletic[1], Nathan Di Pierro[1], Andrew Harrison[1], Linda McIver[1], Sonika Tyagi[2]

1. John Monash Science School, Melbourne, VIC, Australia

2. Monash Bioinformatics Platform, Monash University, Melbourne, VIC, Australia

There is currently no efficient way to accurately identify miRNA or predict the structure of miRNA in an *ab-initio* manner. The purpose of this work is to provide a framework which allows for efficient identification of mature miRNA and folding of pre-miRNA using a feedforward neural network (FFNN) and probabilistic context-free grammar (PCFG) parsing, respectively. The FFNN interprets and provides a prediction of the likelihood, expressed by a probability, of the input being miRNA. The FFNN was trained on a positive set of known human mature miRNAs from the miRBase database and randomly selected sequences from human chromosome 1 used as a negative training set. After training, the FFNN developed an accuracy of 84% when tested on seperate high confidence potential miRNA sequences of which it was not trained on. The probability of a false negative was found to be 16%, while the probability of a false positive when tested on the negative data was found to be 6x10-4%, indicating a high specificity of predicting mature miRNA. The PCFG was created to predict the structures of pre-miRNA, trained on a set of 1800 human pre-miRNA and tested on 400. Out of all control cases using high confidence miRNA, the program returned folded structures that matched the canonical structures to an accuracy of 81%. Further refinement using free energy models could increase this, this would however significantly affect the runtime of the program due to the computationally intensive nature of free energy models. The results of this work indicates definite patterns in miRNA folding and sequences which could facilitate the development and discovery of new strands and their characteristics. Though the current study was done using human data, the probabilistic models are generic and can be trained to work with different organisms.

264

Click chemistry enables preclinical evaluation of targeted epigenetic therapies

Dean S. Tyler[2, 1], Johanna Vappiani[3], Tatiana Cañeque[4], Enid Y.N. Lam[2], Aoife Ward[3], Omer Gilan[2, 1], Yih-Chih Chan[2], Antje Hienzsch[4], Anna Rutkowska[3], Thilo Werner[3], Anne Wagner[3], Dave Lugo[5], Richard Gregory[5], Cesar Ramirez Molina[5], Neil Garton[5], Christopher R. Wellaway[5], Susan Jackson[2], Laura MacPherson[1], Margarida Figueiredo[2], Sabine Stolzenburg[2], Charles C. Bell[2], Colin House[2], Sarah-Jane Dawson[2, 1], Edwin D. Hawkins[6], Gerard Drewes[3], Rab K. Prinjha[5], Raphael Rodriguez[4], Paola Grandi[3], Mark A. Dawson[2, 1]

1. Sir Peter MacCallum Department of Oncology, University of Melbourne, melbourne

2. Petermac, Melbourne, VIC, Australia

3. Cellzome GmbH, GlaxoSmithKline, Heidelberg, Germany

4. Chemical Cell Biology Group, Institut Curie, Paris Sciences et Lettres Research University, Paris, France

5. Epigenetics Discovery Performance Unit, Immuno-Inflammation Therapy Area Unit, GlaxoSmithKline, Stevenage, UK

6. Immunology, Walter and Eliza Hall Institute of Medical Research, Melbourne

The success of new therapies hinges on our ability to understand their molecular and cellular mechanisms of action. We modified BET bromodomain inhibitors, an epigeneticbased therapy, to create functionally conserved compounds that are amenable to click chemistry and can be used as molecular probes in vitro and in vivo. We used click proteomics and click sequencing to explore the gene regulatory function of BRD4 (bromodomain containing protein 4) and the transcriptional changes induced by BET inhibitors. In our studies of mouse models of acute leukemia, we used high-resolution microscopy and flow cytometry to highlight the heterogeneity of drug activity within tumor cells located in different tissue compartments. We also demonstrate the differential distribution and effects of BET inhibitors in normal and malignant cells in vivo. This study provides a potential framework for the preclinical assessment of a wide range of drugs.

265

Characterization of *COL3A1* mutations causing Ehler-Danlos Syndrome type IV: the first step in designing molecular therapies

Sasiwimon Utama[3, 1, 2], May Aung-Htut[2, 4], Sue Fletcher[2, 4], Steve Wilton[2, 4]

1. Biochemistry and Molecular Biology, Mahidol University, Bangkok, Thailand

2. Centre for Comparative Genomics, Perth, WA, Australia

3. Murdoch University, Murdoch, WA, Australia

4. Perron Institute for Neurological and Translational Science, Perth, WA, Australia

Ehler-Danlos syndrome type IV; vascular type (EDS type IV, vEDS), is a dominant disease caused by mutations in *COL3A1*, which encodes collagen III. Collagen III is a major component of collagen fibrils in the extracellular matrix (ECM) of hollow organs such as the lungs, uterus and vascular tissues, as well as skin. Collagen III consists of three identical COL3A1 α chains coiled around each other, a homotriplex that is secreted from cells for fibril assembly. For this reason, COL3A1 α chain defects (missense or premature termination of translation) affect collagen III homotrimer formation and fibril assembly. Mutations in collagen III are reported to affect production, secretion and thermal stability. These biochemical abnormalities influence collagen III ECM assembly leading to vEDS symptoms, which include skin hyper-flexibility, easy bruising, joint hypermobility and delayed wound healing. However, the major symptom that causes death is vascular rupture in the internal organs. The first step in designing any genetic therapy to address vEDS mutations is their precise identification. In this study, we used RNA-based screening to identify mutations in this 51-exon gene. A c.766delA nonsense mutation resulting in a pronounced reduction of mRNA levels due to nonsense mediated decay. Furthermore, splice site mutations, IVS7+5G>A and IVS14-2A>G were detected in patient cells after skipping of exon 7 and exon 15 was identified by RNA-based screening.

Recently, antisense oligonucleotides (AOs) have been approved to by-pass dystrophin protein-truncating mutations in a clinical setting for Duchene muscular dystrophy. Moreover, *COL3A1* has several similarities with *dystrophin*: (i) consists of many exons, some of which can be skipped without affecting function (ii) both encode a long half-life structural protein. According to these similarities, AO mediated splicing could also be a potential therapy for selected *COL3A1* mutations. Exon skipping strategies using AOs to address the nonsense and splice site mutations are described.

266

shRNA screen for novel epigenetic regulations of γ-globin silencing

Jim Vadolas[1], Betty Kao[2], Shian Su[3], Matthew Ritchie[3], Jamie Gearing[3], Miha Pakusch[3], Jovana Maksimovic[2], Alicia Oshlack[2], Marnie Blewitt[3]

1. Hudson Institute of Medical Research, Melbourne, VIC, Australia

2. Murdoch Childrens Research Institute, Parkville, VIC, Australia

3. Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia

A greater understanding of the epigenetic processes that contribute to γ-globin silencing will ultimately help inform the development of therapeutic strategies for the reactivation of γ-globin. In this study, we generated a murine erythroleukaemia cell line that carries an intact 183kb human β-globin locus modified to express eGFP under the control of the γ-globin promoter (GG-MEL). The GG-MEL reporter was used to screen a retroviral shRNA library containing 1027 shRNAs targeting 158 genes encoding epigenetic regulators. Following retroviral transduction, GFP-positive and GFP-negative samples were collected over the course of differentiation. Next generation sequencing and bioinformatics analysis were used to compare differences in shRNA representation between the GFP-positive and GFP-negative samples. The screen identified several genes that are known regulators of γ-globin silencing, as well as genes that have not previously been associated with γ-globin regulation. Following validation of novel candidates, *Kdm2b* emerged as the top regulator of γ-globin silencing.

Molecular characterisation was conducted in parental MEL cells and included RNA-seq to investigate global changes in gene expression upon *Kdm2b* knockdown, and ChIP-seq to identify global binding sites of KDM2B. ChIP-seq analysis revealed strong enrichment of KDM2B at actively expressed β-globin genes, but was absent at the developmentally silenced embryonic/fetal globin genes. Knockdown of *Kdm2b* resulted in increased expression of the γ-globin reporter in primary erythroid cells derived from the GG murine fetal livers. These results suggest *Kdm2b* may be an activator of β-globin gene expression and down regulates γ-globin through a competitive model.

More recently, KDM2B has been identified to be a component of the non-canonical polycomb repressive complex PRC1.1, which is associated with activating histone modifications. We propose that KDM2B regulates

the β-globin gene through interaction with non-canonical PRC1.1 complex and protects the gene from DNA hypermethylation.

267

Novel contribution of acetylated histone variant H2A.Z in activation of neo-enhancers in prostate cancer.

Fatima Valdes Mora[1], Cathryn Gould[1], Yolanda Colino-Sanguino[1], Wenjia Qu[1], Jenny Song[1], Aaron Statham[1], Shalima Nair[1], Nicola Armstrong[2], James Kench[3], Kenneth Lee[4], Lisa Horvath[1], Minru Qiu[5], David Gallego Ortega[1], Clare Stirzaker[1], Susan Clark[1]

1. Garvan Institute of Medical Research, Darlinghurst, NSW, Australia

2. Murdoch University, Perth, WA, Australia

3. Royal Prince Alfred Hospital, Camperdown, NSW, Australia

4. Concord Hospital, Sydney, NSW, Australia

5. Department of Anatomical Pathology, Sydpath,, St Vincent's Hospital, Darlinghurst, NSW, Australia

Acetylation of the histone variant H2A.Z (H2A.Zac) occurs at active promoters and is associated with oncogene activation in prostate cancer, but its role in enhancer function is still poorly understood. Here we show that H2A.Zac containing nucleosomes are commonly redistributed to neo-enhancers in cancer resulting in a concomitant gain of chromatin accessibility and ectopic gene expression. Notably incorporation of acetylated H2A.Z nucleosomes is a pre-requisite for activation of Androgen receptor (AR) associated enhancers. H2A.Zac nucleosome occupancy is rapidly remodelled to flank the AR sites to initiate the formation of nucleosome-free regions and the production of AR-enhancer RNAs upon androgen treatment. Remarkably higher levels of global H2A.Zac correlate with poorer prognosis. Together these data demonstrate the novel contribution of H2A.Zac in activation of newly formed enhancers in prostate cancer.

268

Transcriptional characterization of low input frozen brain samples at single-nucleus resolution using 10x Genomics Chromium microfluidics

Dulce Vargas Landin[1, 2], Rebecca Simmons[1, 2], Marina Oliva[1, 2], Daniel Poppe[1, 2], Jahnvi Pflueger[1, 2], Ryan Lister[1, 2]

1. The University of Western Australia, Crawley, WA, Australia

2. Epigenetics and Genomics Laboratory, Harry Perkins Institute of Medical Research, Perth, Western Australia, Australia

The recent development of single-nucleus RNA-sequencing (snRNA-seq) techniques has exposed the cellular diversity of complex tissues and new insights in cellular differentiation, tissue development, and human diseases. There are two leading snRNA-seq techniques, the DroNc-seq and the 10X Genomics Chromium system, both based on microfluidics to encapsulate single nuclei with unique barcodes that allow nucleus identification. These two techniques are capable of profiling thousands of nuclei and classifying them by cell types. However, they require large amounts of input material and, in the case of 10X Genomics Chromium, fresh tissue. These two disadvantages restrict their usage for clinical samples, which are usually frozen and in limited sizes. Here we present a single-nucleus isolation protocol for low input frozen samples. This protocol is compatible with the 10X Genomics Chromium microfluidics system and can be completed in 40 minutes. Using this protocol, we successfully profiled four frozen human and mouse cortices with 15-20 mg starting material. We compared the data obtained from our frozen samples to the publicly available libraries from fresh tissue generated by the 10X Genomics Chromium. With our technique, we detected a similar number of genes (~1,500), obtained a higher diversity of cell types and had lower mitochondrial transcripts (<0.05%) than in the fresh tissue samples. Overall, this protocol allows the transcriptional characterization of small frozen samples at single nucleus resolution, increasing the clinical applications of snRNA-seq.

1. Habib, N., et al., Massively parallel single-nucleus RNA-seq with DroNc-seq. Nat Methods, 2017. 14(10): p. 955-958.
2. Lake, B.B., et al., A comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA. Sci Rep, 2017. 7(1): p. 6031.

269

Racing against the (epigenetic) clock: physical activity reverses age-related epigenetic drift in human skeletal muscle

Sarah Voisin[1], Nir Eynon[1, 2], Severine Lamon[3], Xu Yan[1], Nicholas Harvey[4], Kevin J Ashton[5], Larisa M Haupt[4], Fiona Munson[1], Andrew Garnham[1], David J Bishop[1], Lyn R Griffiths[4]

1. Victoria University, Melbourne, VIC, Australia

2. MCRI, Melbourne, VIC, Australia

3. School of Exercise & Nut. Sci., Deakin University, Melbourne, VIC, Australia

4. School of Biomedical Sciences, Queensland University of Technology, Brisbane, QLD, Australia

5. Bond Institute of Health and Sport (BIHS), Bond University, Gold Coast, QLD, Australia

Alterations of the epigenetic landscape are a hallmark of aging, leading to aberrant gene expression profiles. Moreover, physical activity induces widespread changes in the methylome and miRNA expression profiles of various tissues, especially skeletal muscle. We hypothesized that 1) higher levels of physical activity are negatively associated with age-related epigenetic drift and 2) a 4-week high-intensity interval training program can reverse age-induced epigenetic drift.

To test our hypotheses, we generated genome-wide DNA methylation and miRNA expression profiles in skeletal muscle from 25 individuals who participated in the Gene Skeletal Muscle Adaptive Response to Training (SMART) study [1]. Participants were all healthy, recreationally active, Caucasian men aged 18-45. Muscle biopsies were taken from the *vastus lateralis* muscle at baseline and at the end of the 4 weeks of training. DNA methylation was assessed with the Illumina EPIC chip and miRNA expression profiles were generated with the TaqMan® Array Human MicroRNA A+B Cards Set v3.0. All preprocessing and downstream data analyses were performed in R with the *limma* package.

We found a strong inverse relationship between age-related and fitness-related epigenetic changes at baseline for both DNA methylation ($r = -0.40$, $p < 0.001$) and miRNA expression ($r = -0.58$, $p < 0.001$). We also found a strong inverse relationship between age-related methylome changes and methylome changes induced by four weeks of high-intensity interval training ($r = -0.21$, $p<0.001$), but this was not the case for miRNA expression ($r = 0.070$, $p=0.35$).

In summary, we found that fitter subjects displayed younger epigenetic profiles in skeletal muscle at baseline and an exercise intervention was able to shift DNA methylation profiles to younger profiles, but this was not the case for miRNA expression. This is the first direct evidence that physical activity can offset the effects of aging at the epigenetic level in skeletal muscle.

1. [1] Yan, X. et al. The Gene SMART study: Method, Study Design, and Preliminary Findings. BMC Genomics (2017).

270

Dissecting the roles of eIF2 and eIF3 during translation using TCP-seq

Susan Wagner[1], Stuart K Archer[2], Nikolay E Shirokikh[1], Neelam Sen[3], Nick T Ingolia[4], Alan G Hinnebusch[3], Thomas Preiss[1, 5], Leos S Valasek[6]

1. The John Curtin School of Medical Research, The Australian National University, Acton, ACT, Australia

2. Monash Bioinformatics Platform, Monash University, Melbourne, Vic, Australia

3. Laboratory of Gene Regulation and Development, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA

4. Department of Molecular and Cell Biology, Center for RNA Systems Biology, University of California, Berkeley, California, USA

5. Victor Chang Cardiac Research Institute, Sydney, NSW, Australia

6. Institute of Microbiology, Czech Academy of Sciences, Prague, The Czech Republic

Regulation of mRNA translation is a key element of eukaryotic gene expression control and, for example, prominently involved in adjusting the proteome to rapid changes in the cellular environment. The most highly regulated phase of translation is the initiation phase, during which ribosomes are recruited to the mRNA 5' untranslated region (UTR). Over a dozen eukaryotic initiation factors (eIFs) assist the small ribosomal subunit (SSU) in binding near the 5' cap, scan the 5' UTR, recognise the appropriate start codon and unite with the large ribosomal subunit (LSU). Two key factors involved in these processes are eIF3, a multifunctional factor that wraps around much of the SSU, and eIF2, which forms a complex with initiator tRNA and is pivotal in start codon recognition. The prevalence of upstream open reading frames and near-cognate start codons among mRNA makes start codon recognition an important target for quantitative and qualitative control of translation.

In the present study, we use translation complex profile sequencing (TCP-seq) in combination with affinity purification of eIF-SSU complexes to investigate the dynamic involvement of eIF2 and eIF3 in different sub-steps of translation. We focus on the regulatory role of upstream open reading frames (uORFs) and the mechanism of resumption of scanning and reinitiation downstream. We also study aspects of start codon recognition that relate to the poorly understood role of the immediate sequence context surrounding efficient start codons– the 'Kozak context'. We identify translation initiation complexes that differ in their composition and functionality during start codon recognition.

271

Molecular Dynamics Modelling of a Variant of Unknown Effect in RAD51D

Matthew Wakefield[1, 2], Mike Kuiper[2], Olga Kondrashova[1], Kristy Shield-Artin[1], Clare Scott[1]

1. Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia

2. Melbourne Bioinformatics, The Unviersity of Melbourne, Parkville, Vic, Australia

High-grade epithelial ovarian carcinomas (OC) containing mutated *BRCA1/2* have homologous recombination defects and are sensitive to poly(ADP-ribose) polymerase inhibitors (PARPi). In a clinical trial of the PARPi rucaparib (ARIEL2 Part 1, Clovis Oncology) a patient was observed with a germline truncating mutation in *RAD51D* (c.770_776del, p.G258Sfs*50) and a secondary mutation (c.770_776delinsA, p.S257_R259delinsK) in a biopsy of a splenic lesion that was progressing on PARPi therapy. Evolutionary analysis and molecular dynamics modelling were used to assess the function of this variant of unknown effect alongside the functional wild-type variant(s). Results indicated that the observed differences in amino acid sequence between the secondary mutation and wild-type RAD51D were unlikely to disrupt normal function and are evolutionarily well tolerated. The secondary mutation (c.770_776delinsA, p.S257_R259delinsK) would likely mirror the function of wild-type RAD51D, thus would restore function and lead to PARPi resistance. This prediction was confirmed by CRISPR directed homology repair introduction of the secondary mutation into a human ovarian cancer cell line, PEO4, which demonstrated a decreased cisplatin and rucaparib sensitivity relative to a PEO4 *RAD51D* knockout. In conclusion, the secondary *RAD51D* mutation (c.770_776delinsA, p.S257_R259delinsK) identified in this lesion most likely contributed to or caused the PARPi resistance and lesion progression.

1. Kondrashova et al 2017 Cancer Discovery DOI:10.1158/2159-8290.CD-17-0419

272

Development and optimization of an efficient method for CRISPR/Cas correction of patient-specific iPSCs

Qi Wang[1], Kristof Wing[1], Alice Pébay[2, 3], Anthony Cook[4], Alex Hewitt[1, 2, 3]

1. Menzies Institute for Medical Research, University of Tasmania, Hobart, Tasmania, Australia

2. Centre for Eye Research Australia, University of Melbourne, Royal Victorian Eye and Ear Hospital, Melbourne, Victoria, Australia

3. Department of Ophthalmology, University of Melbourne, Melbourne, Victoria, Australia

4. Wicking Dementia Research & Education Centre, University of Tasmania, Hobart, Tasmania, Australia

Patient specific induced pluripotent stem cells (iPSCs) can be differentiated into any cell type of the body and harbour disease-specific genetic profiles, making them an ideal platform for modelling inherited diseases. Although combining iPSCs with CRISPR/Cas technology has been championed as attractive strategy in the development of gene-based therapies, given the low frequency of homology directed repair (HDR) pathway, the generation of isogenic gene-corrected iPSCs is generally laborious. Here we report the optimization of a high throughput CRISPR/Cas correction of patient-specific iPSCs. We nucleofect a fluorescent-labelled sgRNA with high fidelity form of Cas9 as a CRISPR ribonucleoprotein (RNP) along with an asymmetric repair donor to target the mutation site of iPSCs. Post-electroporation, the cells are cultured with L755507 for 48h to enhance the HDR efficiency, then rather than manually picking single clones, Fluorescence Assisted Cell Sorting (FACS) is used to seed individual cells into vitronectin coated 96-well plates. Following adaptation to CloneR supplement and a plate centrifuge process, we observed a final single cell viability of at least 50% on feeder-free culture. Screening of single cell clones is initially performed with PCR amplification and single nucleotide polymorphism (SNP) genotyping. Clones are finally screened via bi-directional sanger sequencing. Using this approach, isogenic corrected patient-specific iPSCs can be generated within 5 weeks for a cost of approximately $AUD 3000. The combination of these optimizations has allowed the development of a rapid, easy, and efficient means for gene correction in iPSCs.

273

An optimal approach to assembling chloroplast genome

Weiwen wang[1], Miriam Schalamun, Alejandro Morales-Suarez, David Kainer, Benjamin Schwessinger, Robert Lanfear[1]

1. Research School of Biology, Australian National University, Canberra, ACT, Australia

Properly assembled organelle genomes are vital to fields from medicine to phylogenetics. New long-read sequencers provide opportunities for improving or even perfecting organelle assemblies, but little is known about the best way to combine long- and short-read assemblies to get the best results. We used the chloroplast genome as a test case, because it contains two large inverted repeats which make it difficult to assemble with only short-read data. Oxford Nanopore long-reads can span the whole chloroplast genome, but have a high error rate. We first developed a method for independently assessing genome assembly quality. We then investigated a range of assembly approaches using long-read only, short-read only and hybrid (i.e. combined long- and short-read) methods. In addition, we determined the minimum coverage of long- and short-reads necessary to achieve the best results. We show that hybrid assemblies with at least 20x coverage of long-reads and 40x coverage of short-reads produce the best results, with a single contig spanning the entire chloroplast genome with a very low error rate. Our results inform best-practice for assembling both animal and plant organelle genomes.

274

Characterizing the epigenetic modifier Smchd1 in X chromosome inactivation

Iromi Wanigasuriya[1], Quentin Gouil[1], Andrew Keniry[1], Tamara Mclennan[1], Natasha Jansz[1], Niall Geoghegan[1], Kelly Rogers[1], Tarcy Willson[1], Chen Kelan[1], Graham Kay[2], Kelsey Breslin[1], Matthew E Ritchie[1], Marnie E Blewitt[1]

1. The Walter and Eliza Hall Institution of Medical Research, Parkville, VIC, Australia

2. Queensland Institute of Medical Research, Brisbane, Queensland, Australia

X chromosome inactivation (XCI) is a coordinated set of epigenetic mechanisms that leads to inactivation of the majority of genes on one of the two X chromosomes in female mammals. During XCI, gene silencing happens in three main stages: initiation, establishment and maintenance. Structural maintenance of chromosomes flexible hinge domain containing 1 (*Smchd1*) is an epigenetic modifier that plays an important role in XCI. Although previous studies demonstrate that Smchd1 is important in maintenance of XCI, the role of Smchd1 in establishment is not been clearly defined. Our initial immunofluorescence studies in mouse epiblast stem cells (mEpiSC) derived from E6.5 and E7.5 embryos suggest that Smchd1 is recruited to the inactive X (Xi) between E6.5 to E7.5. To further understand Smchd1's role in XCI during this developmental window, we are performing allele-specific RNA-seq in Smchd1 deleted embryos. To understand the dynamics of Smchd1 binding to the Xi, we performed Fluorescence Recovery After Photo bleaching (FRAP) on immortalised female mouse embryonic fibroblasts (MEFs) carrying a *Smchd1*-GFP knock-in allele. We analysed both Smchd1 wild type and our gain of function Smchd1 mutant (Fresia) and have identified that there is a significant increase in the rate of reloading in the Fresia mutant, indicating differences in Smchd1-GFP dynamics within the nucleus. We plan to carry out similar studies in cells that carry a loss of function mutation in Smchd1. We are also using lattice light sheet microscopy to understand the reloading kinetics of Smchd1 during cell division. Finally, we performed 3D Structural illumination microscopy (3D-SIM) studies to discover possible variation in structural conformation of Smchd1 protein in Xi by tagging both ends of the protein molecule. This will help to further our understanding of the molecular mechanisms of Smchd1 that contribute to both normal gene silencing and disease.

275

NxtIRF: A novel computational approach to measure differential intron retention in cancer databases

Alex CH Wong[2, 1, 3], John JE Rasko[2, 3, 4], Justin J-L Wong[2, 1, 3]

1. Gene Regulation in Cancer Laboratory, Centenary Institute, Camperdown, NSW, Australia

2. Gene and Stem Cell Therapy Program, Centenary Institute, Camperdown, NSW, Australia

3. Sydney Medical School, University of Sydney, Camperdown, NSW, Australia

4. Cell and Molecular Therapies, Royal Prince Alfred Hospital, Camperdown, NSW, Australia

Background: Intron retention (IR), where specific introns are retained in polyadenylated mature messenger RNA, is a mode of alternative splicing that modulates gene expression in normal physiology and cancer, including haematological malignancies. Existing applications (e.g. IRFinder, MISO, rMATS) measure

differential IR but none offer a streamlined approach for IR analyses of high numbers of samples in publicly available RNA-sequencing databases.

Method: We created NxtIRF, a versatile R package designed to streamline differential IR analysis downstream to our existing in-house algorithm, IRFinder[1]. NxtIRF uses an enhanced algorithm to reduce false over-calling of IR, and create data structures to analyse ≤1000 samples. NxtIRF also streamlines data visualization including volcano plots, principal component analysis (PCA), heatmap generation, hierarchical clustering, and annotation track representation of IR events. Additionally, we outline an algorithm that summates transcriptome-wide IR, a measurement used to stratify samples for IR-associated differential gene expression analysis. NxtIRF and IRFinder was used to process polyA-enriched mRNA-seq from acute myeloid leukaemia (AML, n=133) and diffuse large B-cell lymphoma (DLBCL, n=48) samples from The Cancer Genome Atlas. IR levels were compared with unpaired normal bone marrow[2] and lymphoid tissue[3] respectively.

Outcome: Consistent with a previous report[4], IR predominantly increased in AML compared to normal bone marrow. DLBCL exhibits dramatically reduced IR compared to B lymphocytes in normal lymphoid tissue. This observation is striking as breast cancer was the only tumour type previously reported to have decreased IR in tumour versus matched normals[4]. PCA demonstrate DLBCL expresses two distinct IR-expression signatures, suggesting different sets of parent genes are alternately regulated through reduced IR. Distinct genes are differentially expressed in DLBCL and AML when samples are stratified using our transcriptome-wide IR parameter.

Conclusion: NxtIRF facilitates downstream differential IR analysis of large data sets, ideal for bioinformatic analysis of mRNA-seq of large cancer databases to study IR.

1. Middleton R, Gao D, Thomas A, Singh B, Au A, Wong JJ-L, et al. IRFinder: assessing the impact of intron retention on mammalian gene expression. Genome Biol. 2017;18:51.
2. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015;347:1260419.
3. Béguelin W, Popovic R, Teater M, Jiang Y, Bunting KL, Rosen M, et al. EZH2 Is Required for Germinal Center Formation and Somatic EZH2 Mutations Promote Lymphoid Transformation. Cancer Cell. 2013;23:677–92.
4. Dvinge H, Bradley RK. Widespread intron retention diversifies most cancer transcriptomes. Genome Med. 2015;7:45.

276

Roles of histone variant H3.3 and ATRX in tumourigenesis

Maheshi Udugama[1], Hsiao Voon[1], Elaine Sanij[2], Linda Hii[1], Jeremy Henson[3], Jin Bae Son[2], Rick Pearson[2], Philippe Collas[4], Ross Hannan[5], Lee Wong[1]

1. Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria, Australia

2. Peter MacCallum Cancer Centre, Melbourne

3. Prince of Wales Clinical School, Sydney

4. Institute of Basic Medical Sciences, Faculty of Medicine, University of Oslo, Oslo, Norway

5. John Curtin School of Medical Research, Canberra

ATRX (Alpha thalassemia/mental retardation X-linked) complexes with DAXX to deposit histone variant H3.3 into repetitive heterochromatin. A number of studies have shown that specific point mutations in H3.3, ATRX and DAXX are initiating events in a range of cancers including brain cancers, pancreatic neuroendocrine tumours, chondroblastomas and osteosarcomas. In these cancers, ATRX mutations are linked to an Alternative Lengthening of Telomeres (ALT) phenotype. ALT is an aberrant DNA recombination mechanism that drives telomere DNA elongation independent of telomerase activity and is found in 15% of all cancers. Here we investigate chromatin defects associated with H3.3 mutation (a G34R substitution) and ATRX knockout in embryonic stem cells (ESCs).

We introduced a single-copy H3.3 G34R mutation in ESCs. The expression of H3.3 G34R mutant was found to lead to increases in H3K36me3 and H3K9me3 profiles across the genome which were comparable to a KDM4A/B/C triple-knockout. Our study illustrates that histone point mutations can exert a whole-genome effect to promote cancer formation. To study the role of ATRX in tumourigenesis, we knocked out the Atrx gene in ESCs. We showed that depletion of ATRX in mouse embryonic stem cells led to repeat instability and copy loss in ribosomal RNA gene (rDNA). Supporting this, ALT positive human sarcoma tumour samples show a substantially lower rDNA copy than ALT negative samples. Further investigation shows that the rDNA instability is caused by a disruption in H3.3 deposition and thus, a failure in heterochromatin formation in the absence of ATRX. In addition, ATRX-depleted cells are reduced in rRNA transcription output, and show increased sensitivity to RNA polymerase I (Pol I) transcription inhibitor CX5461. Our study provides insights into the

contribution of ATRX loss of function to tumourigenesis through the loss of rDNA stability, and suggests the therapeutic potential of targeting Pol I transcription in ALT cancers.

277

Systematic discovery of coding and noncoding transcriptomic variants in liver cancer

Ji-Hye Choi, Byul A Jee, So Mee Kwon, Min Jae Kim, Sarah Yoon, Gyesoon Yoon, Hyun Goo Woo

Genomic studies on hepatocellular carcinoma have revealed huge molecular heterogeneity of the cancer genome, which were closely associated with different pathobiology cancer progression. To unveil the transcriptomic heterogeneity including coding and noncoding RNAs, we performed profiling of mRNA and micro RNA sequencing from the hepatocellular carcinomas and non-tumoral surrounding tissues. Molecular clustering analysis identified two subtypes with distinct expression of mRNAs and noncoding RNAs, which were significantly associated with patients' clinical outcomes. Differentially mutated genes between the subtypes were also found, which included previously well-known cancer driver genes such as CTNNB1 and NOTCH2. MicroRNA profiling demonstrated the subtypes enriched with the expression of microRNA clusters located in chromosome 14 and 19, which have close interactions with transcriptome aberrations. In addition, long noncoding RNA transcripts and a novel fusion transcript were found, which can promote cancer progression. Our analysis provide a comprehensive landscape of coding and noncoding transcriptome variations in hepatocellular carcinoma, providing new insights on the pathobiology of heterogeneous liver cancer progression.

278

DNA methylation and transcription factor binding

Lu Yang[1], Alister P.W Funnell[1], Richard C.M Pearson[1], Kate G.R Quinlan[1], Merlin Crossley[1]

1. School of Biotechnology and Biomolecular Science, University of New South Wales, Sydney, NSW, Australia

Background: DNA methylation can regulate gene expression by aiding or restricting the recruitment of transcription factors. Global genomic studies have revealed that DNA methylation changes both at CpG islands and non CpG island regions during erythroid development and differentiation. Chromatin occupancy data indicate that many transcription factors involved in erythropoiesis contain CpG within their core consensus DNA binding motifs. We hypothesise that methylation of the CpGs within these binding motifs directly influences the binding of erythropoiesis related transcription factors and thus gene regulation.

Results: Using *in vitro* assays we revealed that DNA methylation can either positively or negatively affect binding of important erythroid transcription factors to their DNA target sites. Based on these *in vitro* results, we analysed ChIP-Seq and genome-wide bisulfite sequencing datasets to find target genes of key erythroid transcription factors *in vivo* where DNA methylation status change significantly during erythroid differentiation. We are now validating these *in vivo* target genes using CRISPR/Cas9 mouse models to uncover how target site methylation modulates genetic regulation during erythroid differentiation.

Conclusion: We have shown that DNA methylation at transcription factor binding sites influences the binding affinity of key erythroid transcription factors *in vitro* and are extending these findings to *in vivo* target genes. Our study will improve the understanding of the relationship between DNA methylation, transcription factor binding and gene regulation in erythropoiesis.

279

ZNF827 - a molecular target for telomere maintenance in cancer

Sile Fiona Yang[1], Alexander P Sobinoff[1], Hilda A Pickett[1]

1. Telomere Length Regulation Group, Children's Medical Research Institute , Westmead, NSW, Australia

Telomeres are nucleoprotein structures at the ends of human chromosomes that function to maintain genome stability. The unlimited proliferative capacity of cancer cells is dependent on their acquisition of a telomere maintenance mechanism to counteract the gradual telomere attrition that accompanies cellular proliferation. Alternative Lengthening of Telomeres (ALT) is a homologous recombination (HR)–mediated telomere maintenance mechanism utilised by approximately 15% of cancers in which ALT status often predicts poor prognosis. We have demonstrated that ZNF827 is recruited specifically to ALT telomeres, where it plays a multifaceted role in promoting ALT activity in concert with the nucleosome remodelling and deacetylase (NuRD) complex. Depletion of ZNF827 results in acute loss of ALT cell viability. To date, there are no other known functions of ZNF827, which makes it a novel potential target for ALT inhibition in cancers. Here we

investigate the normal functions of ZNF827, and the mechanistic details of its recruitment and binding to ALT telomeres. We show that ZNF827 directly interacts with telomeric DNA displaying preferential binding to single stranded DNA. ZNF827 colocalises with DNA damage and repair proteins, indicative of a role for ZNF827 in DNA damage and HR-mediated repair. We propose that ZNF827 is a novel molecular target for ALT cancers, and aim to delineate the molecular interactions involved in the DNA damage response at ALT telomeres. Finally, we aim to explore potential synergistic effects of ZNF827 inhibition and DNA damage-inducing anticancer drugs such as topoisomerase inhibitors and PARP inhibitors for improved cancer therapeutics.

280

Identification and characterization of novel cell populations using single-cell RNA-seq: an example on breast cancer T cell infiltrate

Chengzhong Ye[1], Agus Salim[1, 2], Peter Savas[3], Sherene Loi[3], Terence P Speed[1, 4]

1. The Walter and Eliza Hall Institute of Medical Research, Parkville, VICTORIA, Australia

2. Department of Mathematics and Statistics, La Trobe University, Melbourne, VIC, Australia

3. Peter MacCallum Cancer Centre, Melbourne, VIC, Australia

4. Department of Mathematics and Statistics, University of Melbourne, Parkville, VIC, Australia

Single-cell RNA-seq (scRNA-seq) has provided researchers with an unprecedented opportunity to investigate the heterogeneities of cell populations. Recently popularized droplet-based technologies greatly increased the number of cells that can be profiled in each experiment. However, this type of data suffers from even higher technical noise and dropout rate, presenting new computational and analytical challenges. Here we try to identify and characterize a particular group of cells, tissue-resident memory T cells (TRM), in a breast cancer infiltrating T cell population sequenced with *10X Genomics* droplet-based scRNA-seq platform. We performed a combined analysis including clustering, differential expression and trajectory analysis. Particularly we made vital use of imputation to overcome the high dropout rate of the data. A newly developed model, *DECENT*, was used to perform differential expression analysis on the imputed pre-dropout data. We identified 10 unique clusters with distinct gene expression profiles. One cluster showed enrichment for a list of TRM markers and the following differential expression analysis showed concordance with bulk RNA-seq experiments, suggesting that we were able to identify the TRM population unbiasedly with gene expression measurement. We also discovered a small cluster enriched in both TRM markers and mitotic genes, showing that TRM is undergoing active cell division and therefore potentially an active population in breast cancer immunosurveillance.

1. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell, 161(5), 1202-1214 (2015)
2. Risso, D., Perraudeau, F., Gribkova, S., Dudoit, S., & Vert, J. P. ZINB-WaVE: A general and flexible method for signal extraction from single-cell RNA-seq data. bioRxiv, 125112. (2017)
3. Huang, M. et al. Gene expression recovery for single cell RNA sequencing. bioRxiv, 138677. (2017)
4. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nature biotechnology, 32(4), 381-386 (2014)

281

The FUBP1 *Drosophila* ortholog Psi interacts with the Mediator Complex to regulate growth through modulating expression of *MYC* and RNA processing machinery

Olga Zaytseva[2, 1], Linna Guo[2, 1], Naomi Mitchell[1], Maurits Evers[1], Ross Hannan[3, 1], David Levens[4], Leonie Quinn[1]

1. John Curtin School of Medical Research, Canberra, ACT, Australia

2. University of Melbourne, Parkville, VIC, Australia

3. Peter MacCallum Cancer Centre, Melbourne, VIC, Australia

4. National Institute of Health, Bethesda, Maryland, USA

Mammalian FUBP-family proteins play roles in both RNA processing and transcription, by binding single stranded nucleic acids via their KH domains. In *Drosophila*, the three mammalian FUBP proteins are represented by one ortholog, Psi. We have taken advantage of this reduced functional redundancy to demonstrate that an essential *in vivo* function of Psi is control of cell and tissue growth. Analysis of published Co-IP-mass spectrometry screens positioned Psi in an interactome predominantly comprised of RNA Polymerase II transcriptional machinery. Of great interest was the interaction with most subunits of the transcriptional Mediator (MED) complex, a known sensor of developmental signaling inputs. Moreover, manipulation of MED activity modified Psi-dependent growth, which suggests that Psi interacts with the MED complex to integrate developmental growth signals with transcription of growth regulators. Our work indicates

that a key target of the Psi/MED network that impacts tissue growth is the MYC transcription factor. Further to this, transcriptome-wide expression analysis comparing Psi and MYC depleted wing imaginal discs suggests that in addition to MYC-dependent targets, Psi regulates expression of RNA processing machinery. Our future studies are directed towards identifying direct transcriptional targets of Psi.