AMATA

High-Throughput Genomic Technologies October 18th-21st **2009**

AMATA Executive Committee

President:	Assoc Prof Sean Grimmond	Institute of Molecular Bioscience, UQ
Vice-President:	Assoc Prof Gordon Smyth	Walter & Eliza Hall Institute, Melbourne
Treasurer:	Assoc Prof Nigel Swanson	University of Western Australia
Secretary:	Dr Andrew Holloway	Peter MacCallum Cancer Centre
	Dr Ruby C Y Lin	Ramaciotti Centre, UNSW
	Ms Vikki Marshall	Peter MacCallum Cancer Centre
	Assoc Prof Rik Thompson	St Vincent's Institute, Melbourne
	Mr Les McNoe	Otago University
	Mr Geoff Kornfeld	University of New South Wales
	Prof Nick Hayward	Queensland Institute for Medical Research
	Mr Mark van der Hoek	Hanson Institute, Adelaide
	Dr Daniel Catchpoole	The Children's Hospital at Westmead
	Ms Stephanie Palmer	Australian National University

Organising Committee

Chair:	Professor Ian Dawes	University of NSW
Convenor:	Mr Geoff Kornfeld	University of NSW
Trade Liaison:	Dr Helen Speirs	Ramaciotti Centre, UNSW
Hotel Liason:	Dr Ruby Lin	Ramaciotti Centre, UNSW
	Dr Nikola Bowden	University of Newcastle
	Dr Dan Catchpoole	The Children's Hospital at Westmead
	Dr Sonia Catley	Sydney Bioinformatics
	Dr Vanessa Hayes	Childrens Cancer Institute Australia
	Dr Warren Kaplan	Garvan Research Institute
	Dr Rohan Williams	Australian National University
	Dr Alan Wilton	University of NSW

Conference Management:	Mrs Noelene Gordon	CVBS

AMATA 2009

High-Throughput Genomic Technologies

October 18th-21st 2009

The 9th Annual Meeting of the Australasian Microarray & Associated Technologies Association (AMATA)

MAJOR SPONSORS

The AMATA 2009 Organising Committee gratefully acknowledges the financial support of our sponsors:

PLATINUM SPONSORS:

millennium science

Affymetrix

illumina

Roche

ASBMB SPONSORSHIP:

ASBMB Australian Society for Biochemistry and Molecular Biology

GOLD SPONSORS:

Applied Biosystems

PLP

NCRIS 5.1 NSW

SILVER SPONSOR:

agrf Australian Genome Research Facility Ltd

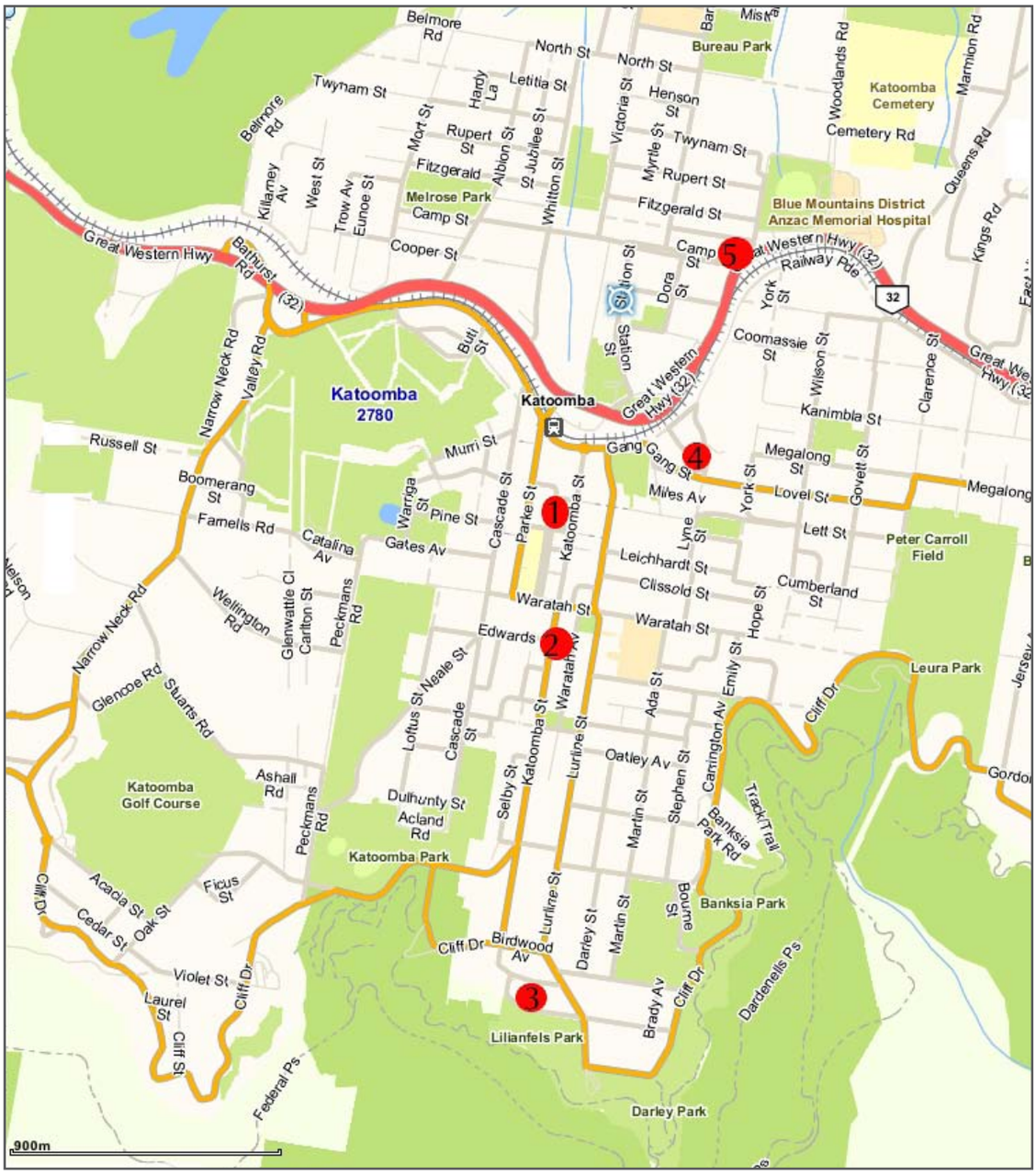
WORKSHOP SPONSORS:

The Ramaciotti Centre for Gene Function Analysis

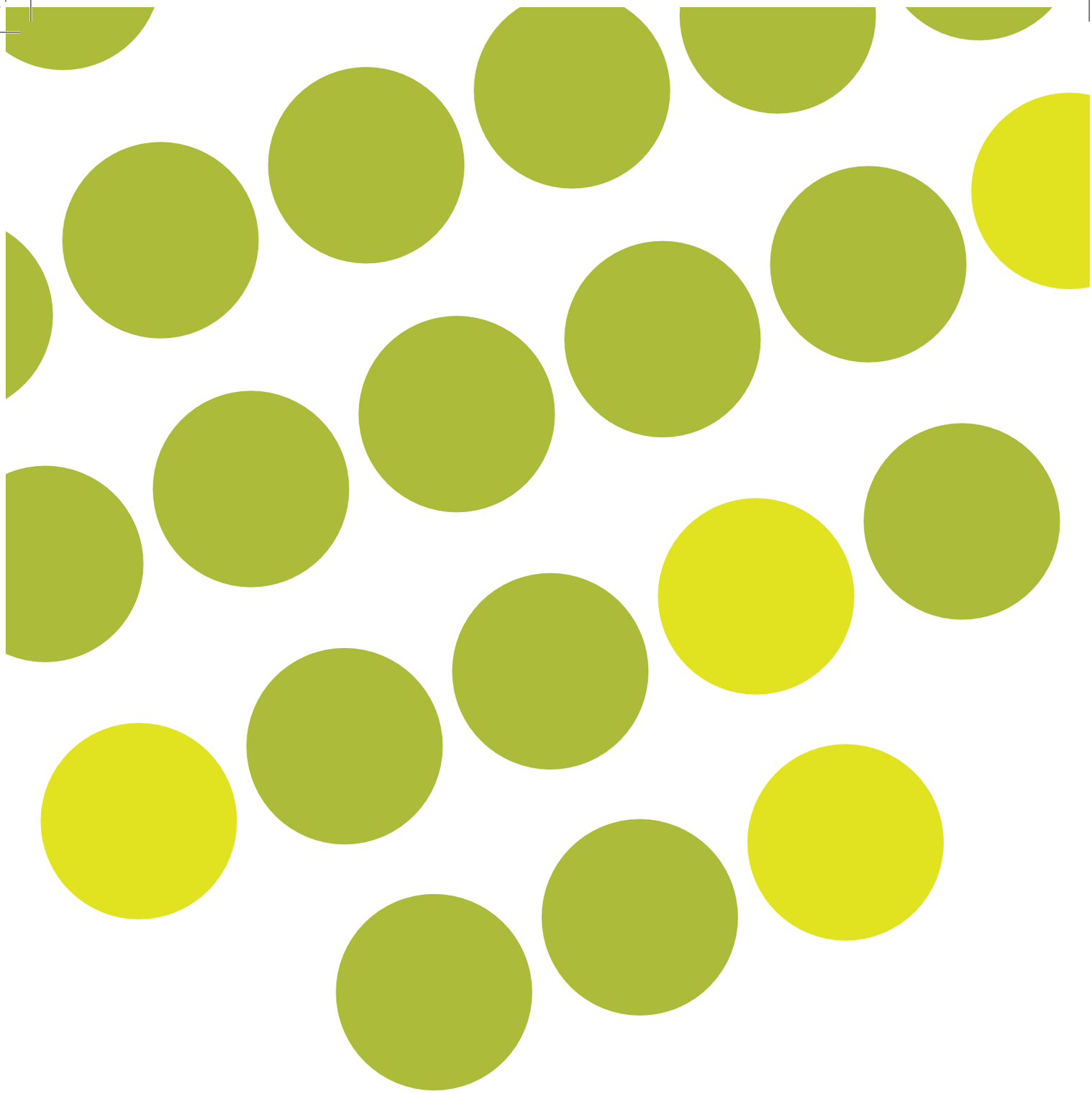
UNSW

kids research Institute the children's hospital at Westmead Tumour Bank

MAP OF KATOOMBA



1. **Carrington Hotel**
2. **Palais Royale**
3. **Lilianfels**
4. **Mountain Heritage**
5. **Alpine Best Western Motel**



PROGRAM

AMATA 2009 PROGRAM

SUNDAY 18th October

10:50	Bus departs from Sydney Airport
11:00 – 14:00	Registration opens

Keynote	Plenary	Guest	Trade	Student

SESSION 1 Variation Genomic — Chair: Rohan Williams

13:45 - 14:00	Welcome	
14:00 - 14:50	Greg Gibson	Geographical Genomics, Canalization, and the origins of human disease
14:50 - 15:20	Carsten Kulheim	Combining next generation high throughput sequencing and population genomics
15:20 – 15:40	Peter Bundock	Discovery of single nucleotide polymorphisms in sugarcane for gene mapping using 454 sequencing
15:40 – 16:00	Vicky Cho	Measuring inter-individual variation in gene expression in monoamine neurotransmitter pathway genes
16:00 – 16:30	Afternoon Tea	

SESSION 2 Genomic Sequencing — Chair: Alan Wilton

16:30 – 17:20	Stephan Schuster	Genomics of extinct and endangered species
17:20 – 17:50	Izhak Haviv	Massively parallel sequencing of epigenomic and functional genomic landscapes
17:50 – 18:10	Petra Souter	The transcriptome of an adult coral and its symbiotic zooxanthellae characterized using 454 pyrosequencing
18:10 – 18:30	Jeremy Shearman	Identifying cause of Cerebellar Abiotrophy in Australian Kelpie dogs using SNP arrays, Nimblegen sequence capture arrays and 454 sequencing.
18:30 – 19:30	Mixer	
TBA	Organising Committee Dinner with Invited Speakers	

MONDAY 19th October

7:00	Exhibitor Setup in Ballroom	
7:30	Registration desk open	

SESSION 3 Transcriptional Complexity — Chair: Daniel Catchpoole

8:30 – 8:50	Anthony Beckhouse	Widespread alternate transcription in human monocyte derived macrophages
8:50 – 9:20	Marcel Dinger	Beyond the exome: Re-interpretation of genome-wide association studies using an empirically-based annotation of the noncoding transcriptome
9:20 – 9:40	Nicole Cloonan	Sequencing the transcriptional complexity of small rnas
9:40 – 10:30	Patrick Brown	The dark matter of biological regulation
10:30 – 11:00	Morning Tea	

SESSION 4 Regulation of Transcription — Chair: Jean Yang

11:00 – 11:50	Martha Bulyk	High-resolution DNA binding specificity profiles of transcription factors and cis regulatory codes in DNA
11:50 – 12:20	Richard Pearson	The role of basic krüppel-like factor (Klf3) in b-lymphocyte development and tumorigenesis
12:20 – 12:40	William Ritchie	Bioinformatics of high throughput RNA data
12:40 – 13:00	Oscar Junhong Luo	Systematic discovery of gene regulation networks from genome-wide transcription factor binding motif data
13:00 – 14:00	Lunch	

SESSION 5 Emerging Mechanisms of Transcriptional Regulation — Chair: Ruby Lin

14:00 – 14:50	Muller Fabbri	Causes and consequences of microRNA dysregulation in cancer
14:50 – 15:10	David Humphreys	Small RNA profiling of the HL-1 cardiomyocyte cell line using next generation sequencing
15:10 – 15:30	Jean Yang	Identification of miRNA regulatory modules and their targets using matched miRNA-mRNA data
15:30 – 15:50	Lorey Smith	A large scale RNAi screen for regulators of scribble function in cancer reveals the mammalian cell polarity network as a novel regulator of Ras oncogenic function
15:50 – 16:10	Gavin Huttley	The influence of epigenetic state on genetic variation
16:10 – 16:40	Afternoon Tea	

SESSION 6 Gene Expression Profiling — Chair: Rick Thompson

16:40 – 17:00	Caroline Kerr	Comparative global gene expression patterns of colonic epithelium in azoxymethane-induced carcinogenesis in rats and colorectal cancer in humans
17:00 – 17:20	Briony Jack	Transcription factor GATA2 in the repression of adipogenesis
17:20 – 17:40	Matthew Wakefield	RNA-seq enables whole genome allele specific expression

17:45–19:45	Poster Session

TBA	Pre-AGM AMATA Executive Meeting

TUESDAY 20th October

SESSION 7 Translational Research — Chair: Nikola Bowden

8:30 – 8:50	Guoji Guo	Single cell quantitative gene expression analysis from zygote to blastocyst
8:50 – 9:20	John Pimanda	Differential regulation of tissue specific gene expression by modular assembly of enhancers
9:20 – 9:40	Geoff Morton	Data mining the genetics of leukemia
9:40 – 10:30	Javed Khan	Integrated analysis of the cancer genome

10:30 – 11:00	Morning Tea

SESSION 8 Genome-wide Analysis — Chair: Sean Grimmond

11:00 – 11:50	Vishy Iyer	Analyzing genome-wide transcription and chromatin with microarrays and next-generation sequencing
11:50 – 12:20	Xiaoan Ruan	Chia-PET, A whole-genome approach to study long-range chromatin interactions in mammalian cells
12:20 – 12:40	Nadia Whitelaw	Ppie: a modifier of epigenetic reprogramming in the mouse identified by sequence capture and 454 Sequencing.
12:40 – 13:00	Tom Whitington	Chip-seq chromatin modification data facilitates accurate prediction of tissue-specific transcription factor binding sites
13:00 – 13:20	Alister Funnell	Investigating the erythroid role of the transcription factor Klf3

13:45 – 14:30	Martin Goldberg	Affymetrix' Next Generation Automated Genotyping Platform

13:20 – 14:20	Lunch

14:20 – 19:00	Free Time	

19:00 – 21:00	Conference Dinner	

Wednesday 21st October

SESSION 9 Understanding Complex Systems — Chair: Warren Kaplan/ Mark Cowley

8:30 – 8:50	Belinda Phipson	Genuine association of gene expression profiles
8:50 – 9:20	Romaric Bouveret	Building transcriptional regulatory networks for heart development
9:20 – 9:40	David Fung	Modeling the molecular cell system of hepatocellular carcinoma as an actor-semiotic network
9:40 – 10:30	Olga Troyanskaya	Understanding tissue-specific gene expression and disease from large microarray compendia

10:30 – 11:00	Morning Tea and AMATA AGM	

SESSION 10 Functional Genomics — Chair: Ian Dawes

11:00 – 11:50	David Gresham	Genetic control of growth and survival
11:50 – 12:20	Michal Janitz	Integrated analysis of genomic regulatory regions using cell arrays and next-generation sequencing
12:20 – 12:40	Thomas Gonda	Development of high-throughput, high content, functional screening: ARVEC
12:40 – 13:00	Anthony Bourneman	Characterisation of intra-specific genomic diversity in industrial microorganisms by whole-genome next-generation sequencing
13:00 – 13:15		Conference Close: Concluding remarks

13:30	Bus departs for Sydney Airport	

13:30 – 14:30	Lunch	

Workshops

Wednesday 21st October

WORKSHOP 1

14:30 – 18:30	Library	Microarrays & Next-Gen Sequencing in Undergraduate Education

WORKSHOP 2

14:30 – 18:30	Grand Dining Room	Next-Gen Sequencing Workshop
16:10 – 16:40	Afternoon Tea	

Thursday 22nd October

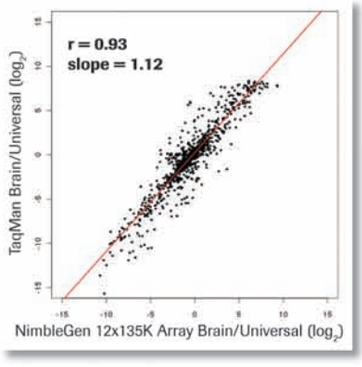
WORKSHOP 3

9:00 – 13:00	Grand Dining Room	AGRF Next-Gen Sequencing Workshop

WORKSHOP 4

9:00 – 13:00	Contact Dan Catchpoole	Virtual Microscopy
10:30 – 11:00	Morning Tea	

Unleash the full potential of NimbleGen Gene Expression arrays...



High-quality Gene Expression data using the NimbleGen MS 200 Microarray Scanner. Highly accurate gene expression profiles achieved with NimbleGen Human Gene Expression 12x135K array and the MS 200 scanner are evidenced by the high concordance with TaqMan data (MAQC report, *Nature Biotechnology*, 2006, 24:1151). Fold-change analysis of 961 genes in brain and universal reference RNA shows a correlation of 0.93 with a slope of 1.12. Highly precise gene expression values are achieved using the MS 200 scanner with a CV median of less than 5% compared to 10% with a competitor's scanner.

Introducing the NimbleGen MS 200 Microarray Scanner

The NimbleGen MS 200 reveals the complete picture of high-density NimbleGen Gene Expression arrays with powerful scanning resolution down to 2µm and increased sensitivity. Scan with certainty using the MS 200 to achieve high-quality data from NimbleGen 12x135K multiplex arrays while ensuring compatibility with the next-generation arrays of tomorrow.

- **Enhanced Data Quality:** High-interarray reproducibility with the advanced autofocus and autogain features ensures optimal results.
- **Maximum Throughput:** 48-slide autoloader processes up to 576 samples in a single run using multiplex arrays.
- **Integrated Workflow:** Optimized for NimbleGen Gene Expression arrays, as well as all current and next-generation NimbleGen arrays for CGH, ChIP-chip, and DNA Methylation applications.

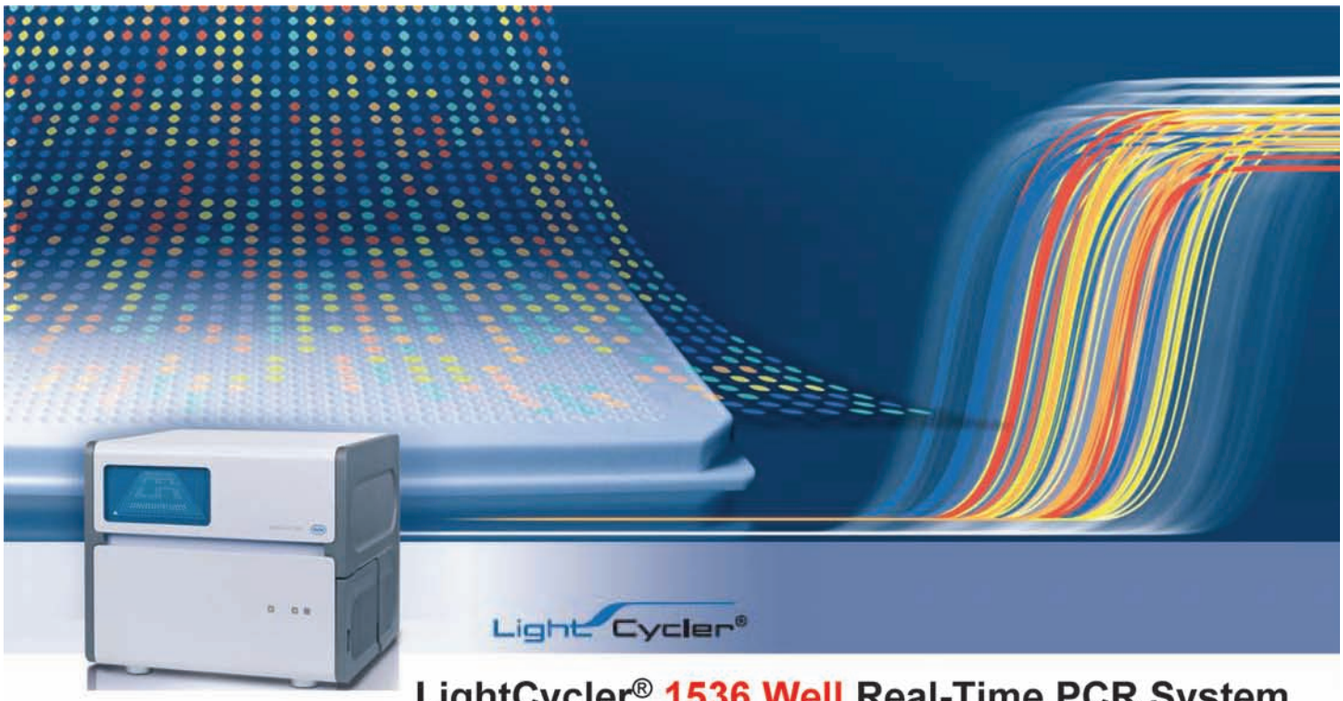
The full potential of NimbleGen Gene Expression arrays is revealed with the MS 200…for fast scans, high-quality data, and consistent research results.

See the NimbleGen MS 200 difference at
www.nimblegen.com/ms200

NimbleGen

Roche NimbleGen, Inc.
Madison, WI USA

Roche



LightCycler® 1536 Well Real-Time PCR System

"Real" High Throughput Built On Proven Technology From The Innovation Leader In Real-Time PCR

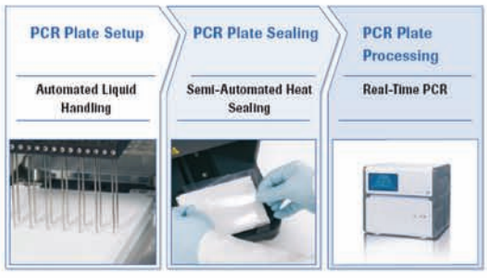


Figure 1: Workflow overview

Figure 2: LightCycler® Multiwell Plates Cutting-edge multiwell plate design for attaining fast, exceptionally precise 1536 real-time PCR data points.

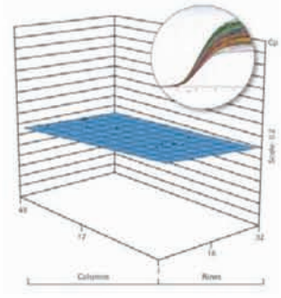


Figure 3: High-performance sensitivity and reproducibility of the LightCycler® 1536 System.

- **Generate 1536 Data Points** in a single run in less than 50 minutes
- Complete **Assay Flexibility**
- **Assay Miniaturization** - Cost reduction per sample point and reduced sample input
- **High Dynamic Detection/sensitivity Range** and **Reproducibility** for cross platform validation
- **Enhanced Software functions** offering powerful capabilities
- **Faster PCR Protocols** with high concentration, long-term stable Real-Time Ready® masters
- **Quality assured and monitored assay set-up system**

Roche Diagnostics
Australia Pty limited
ABN 29 003 001 205
31 Victoria Ave
Castle Hill NSW 2154
Australia
Ph: +61-02-9860 2222
Fax: +61-02-9860 2111

Roche

Posters

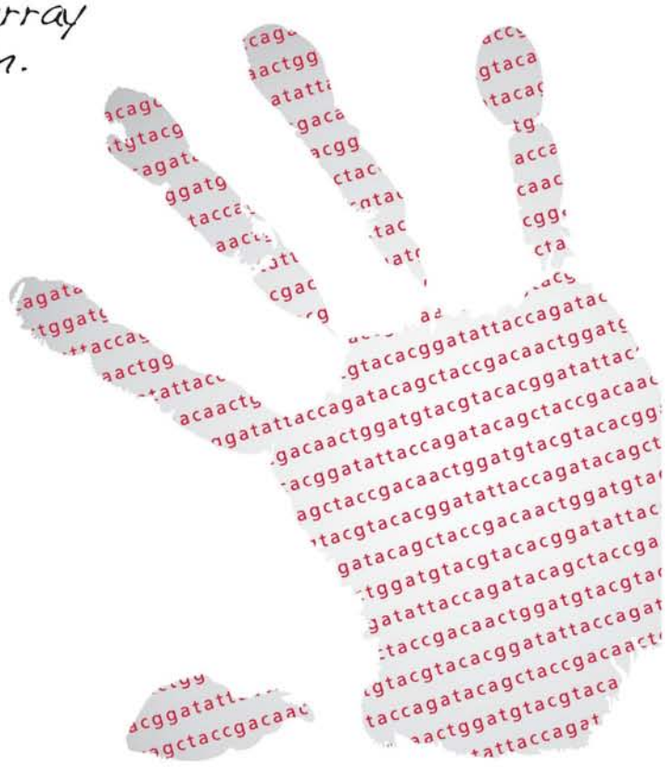
1	**Matthew J. Anderson**, Amanda L. Miotto, Nick A. Matigian, Alan Mackay-Sim, Alistair M. Chalk, Christine A. Wells	**Transcriptional Framework and Expression Viewer: A database for integrating multiple gene expression platforms with transcriptional annotations**
2	Carla F. Kairupan, Nikola A. Bowden, **Katie A. Ashton**, Xu Zhang, Peter Hersey, Rodney J. Scott	**Gene expression profiling in malignant melanoma**
3	**Timothy L. Bailey**, Fabian Buske, Denis C. Bauer, and Mikael Boden	**Assigning roles to regulatory motifs using comparative genomics**
4	**Katherine J Baines**, Jodie L Simpson, Rodney J Scott, Lisa G Wood and Peter G Gibson	**Genome wide gene expression of induced sputum in non-eosinophilic asthma**
5	**Traude Beilharz**, Jeff Squires, Meghna Sobti, Arthur Liu, David Humphreys, and Thomas Preiss	**Analysing post-transcriptional gene control by next generation sequencing**
6	**Tony Blick**, Cletus A Pinto, Mark Waltham and Erik W Thompson	**Pharmacogenomic approaches to the NCI60 database for breast cancer therapeutics**
7	**Nikola A. Bowden**, Katie A. Ashton, Geoffrey Stibbard, Mathew B. Cox, Katherine Baines and Rodney J. Scott	**Predicting *Xeroderma Pigmentosum* complementation group by gene expression profiling**
8	**Lauren M. Bragg**, Stuart E. Denman, Paul Evans, Glenn Stone, Jagadish Padmanabha, Andre D. Wright, Chris S. McSweeney, Dave Edwards, and Mark Morrison	**Challenges of assembling next-generation sequencing data generated from microbial enrichments**
9	**Albert Chetcuti**, Nicole Mackie, Kerrie Jones, Guy Nelmes and Daniel Catchpoole	**The Children's Hospital at Westmead Tumour Bank: A dynamic biospecimen resource for translational cancer research**
10	**Albert Chetcuti**, Safiye Aktas, Nicole Mackie, Celal Ulger, Gokce Toruner, Joseph Rinaggio, Anthony Galante, Bin Tian, Patricia Soteropoulos, Meera Hameed, Jane Carpenter, Marvin Schwalb, James Dermody, Mualla Alkan, and Daniel Catchpoole	**Identification of new proteins to determine invasion in favourable Wilms tumour**
11	**Joyce Chiu**, Carole M. Tactacan, Ruby C.Y. Lin, Merridee A. Wouters and Ian W. Dawes	**Oxidative stress sensing in Saccharomyces cerevisiae by the cell cycle transcription factor Swi6p**
12	**Mathew B. Cox**, Nikola A. Bowden, Rodney J. Scott, and Jeannette Lechner-Scott	**Gene expression profiling in Multiple Sclerosis**
13	**CA de Graaf**, J Choi, T Baldwin, Sargeant TJ, Robinson A, Smyth GK and Hilton DJ	**The Hematopoietic Transcriptome Atlas**
14	**Ken Dutton-Regester**, Lauren Aoude,. Chris Scmidt, Adrian Herington, and Nick Hayward	**High throughput oncogene mutation screening in melanoma cell lines**
15	**WMH d'Avigdor**, M Stapelberg, GW McCaughan, GW McCaughan, Maggie Lee, FJ Warner, and NA Shackel	**Identification of miRNA expression profiles differentiating hepatitis c genotype 1 and 3 liver injury**
16	**Tiffany-Jane Evans**, Nikola A. Bowden, Bente A. Talseth-Palmer, Daniel Catchpoole, and Rodney J. Scott	**Copy number variation in childhood Acute Lymphoblastic Leukaemia**
17	**Sam Forster**, Shamith Samarajiwa, Katie Auchettl, Jodee Gould and Paul Hertzog	**Interferome: Integration of microarray datasets in the search for biological targets**
18	**Hugh J French**, Kaiman Peng, Kristine Hardy, M Frances Shannon, and Rohan BH Williams	**Detecting changes in chromatin state over time.**
19	**Evgeny A. Glazov**, Liang Zhao, Paul Leo, Ping Zhang, Diwakar R. Pattabiraman, Matthew A. Brown, and Thomas J. Gonda	**Defining MYB transcriptional network**
20	**Apurv Goel**, Simone Li, Chi Nam Ignatius Pang, Marc R. Wilkins	**Visualising time-series microarray data on the interactome**
21	**Robin Anderson**, Wen Qiu, Ian Campbell, Adam Kowalczyk, Christina Restall Richard Tothill, Alex Boussioutas, David Bowtell, Terry Speed, Min Hu, and Kornelia Polyak	**Where do carcinoma-associated fibroblasts come from?**
22	**Jennifer Henderson**, Vicky Cho, Laura Garvican, Bon Gray, Chris Gore, Rohan Williams, and Simon Easteal	**Altered transcriptional responses in elite endurance athletes exposed to moderate altitude**
23	**Joshua W.K. Ho** and Michael A. Charleston	**Differential variability of gene expression in human diseases and ageing**
24	**Y. Hu** and G. K. Smyth	**Gene expression analysis of colorectal cancer subtypes based on the Jass system**

25	**K.Krishnan,** N.Cloonan, C.Wong, A.Moller, and S.Grimmond	**Next-generation sequencing of the molecular events in breast cancer**
26	**Charity Law**, Catherine Carmichael, Wei Shi, and Gordon K Smyth	**Bioinformatic evidence for transcription factor interactions: AML1 with ERG and ETS2**
27	**David Lovell**, Warren Müller, Jen Taylor, and Alec Zwart	**Caution! Compositions! Can constraints on omics data lead analyses astray?**
28	**Victoria Lyons**, Gal Winter, Patrick J O'Doherty, Ming J Wu, and Vincent J Higgins	**Biological replicate analysis of one-colour Affymetrix microarrays**
29	**Ruth N. MacKinnon** and Lynda J. Campbell	**Independent amplification of 20q11.21: support for an oncogene in a 250kb region in acute myeloid malignancy with deletion of 20q12**
30	**Nicholas Matigian**, Greger Abrahamsen, Ratneswary Sutharsan, Anthony L Cook, Amanda Nouwens, Bernadette Bellette, Alejandra M Vitale, Jiyuan An, Matthew Anderson, Anthony G. Beckhouse, Alistair M Chalk, Julie Cochrane, John J. McGrath, Jyothy Raju, Greg T. Sutherland, Stephen Mahler, George D. Mellick, Stephen A Wood, Carolyn M. Sue, Christine A. Wells, Alan Mackay-Sim	**Disease-specific, neurosphere-derived cells as models for brain disorders**
31	**Patrick J. O'Doherty,** Ryan A. Hyland, Victoria Lyons, Ming J. Wu, and Vincent J. Higgins	**The impact of technical replicates on the interpretation of single-channel Affymetrix microarrays**
32	**V.M. Perreau**, O. Skibina, Y. Hu, M. Binder, J. Field, H. Butzkueven, G.K. Smyth, and T. J. Kilpatrick	**Application of gene set testing to probe regulatory T cell activity in gene expression data from early multiple sclerosis patients.**
33	**Greg Peters**, Artur Darmanian, Dorothy Hung	**CGH microarray as a diagnostic tool in routine testing for genomic disorders: problems and prospects.**
34	**B.F. Piraino**, A.R. Lloyd, U. Vollmer-Conna, on behalf of The Dubbo Infection Outcomes Study investigators.	**SNP-MaP of the severity and duration of acute illness.**
35	**Janet Shaw**, Caroline A. Kerr, Julie Clarke, Trevor Lockett, Richard Head	**Bioinformatic analysis of merged gene expression microarray datasets in rat models of colorectal cancer.**
36	**Chintanu Kumar Sarmah**, Sandhya Samarasinghe, Don Kulasiri, and Daniel Catchpoole	**Cross platform integration of microarray data**
37	**Wei Shi** and Gordon K Smyth	**Estimating the proportion of microarray probes expressed in an RNA sample**
38	**Peter Simpson**, Nic Waddell, Ana-Cristina Vargas, Leo Da Silv,a Juliet French, Kate Hardie, Lynne Reid, Janani Jayanthan, and Sunil Lakhani	**Molecular profiling of archival breast cancer samples using DASL technology.**
39	**Mitchell S Stark**, Sonika Tyagi, Derek Nancarrow, Glen Boyle, Richard A. Sturm and Nicholas K Hayward	**Characterization of the melanoma miRNAome by deep sequencing**
40	**Rathi D Thiagarajan**, Sean M Grimmond, Kylie Georgas, Bree Rumballe, Dave Tang and Melissa H Little	**Identification of compartment-specific anchor genes for key structures within the developing mouse kidney and transcription factor networks**
41	**Jocelyn van den Bergen**, Denise Miles, Andrew Sinclair, and Patrick Western	**Expression profiling of fetal germ cell differentiation**
42	**Paula M. Vaz**, Richard P. Grant, Fionna E. Loughlin, and Joel P. Mackay	**Exon arrays provide insight into the function of ZRANB2**
43	**Logan C Walker,** kConFab, and Amanda B Spurdle	**Investigation of copy number variants and familial breast cancer risk, and their interaction with rare BRCA1/2 sequence variants**
44	Janice R. Aldrich-Wright, **Shaoyu Wang**, and Vincent J. Higgins	**Microarray analysis of a novel platinum drug suggests a potential for an effective treatment against cisplatin resistant tumours**
45	**Helen Williams**, Merlin Crossley, and Kim Bell-Anderson	**Microarray analysis of epididymal white adipose tissue in Klf3 -/- mice**
46	**Auda Eltahla**, Zi Cong Wu, Sven Warris, Webb Miller, Stephan Schuster, and Alan Wilton	**Development of SNP markers for differentiating dingoes from dogs and hybrids.**
47	**Di Wu** and Gordon Smyth	**Patterns across data sets: finding the cell origin of basal-like breast tumours**
48	**Matthew D Young**, Matthew J Wakefield, Gordon K Smyth, and Alicia Oshlack	**Gene ontology testing for RNA-seq: accounting for selection bias**
49	Abd Rahman, et al.	**Next generation sequencing in the undergraduate laboratory: Sequence analysis of Wollemi Pine nuclear and chloroplast genomes**
50	**Alan Wilton**, Brian McEvoy, Peter Visscher, Eric Wang, Robert Moyzis, Sheila van Holst Pellekaan	**SNP genotyping of samples from Riverine with mixed Australian indigenous and non-indigenous ancestry**

Affymetrix®

Revolutionize life

The world's first hands-free, automated microarray processing system.



GeneTitan™ System

Less time acquiring data, more time for science.

To Learn more, visit the Affymetrix/ Millennium Science booth

millennium science 10 YEARS 1999 - 2009

millennium science

enabling next-generation research

Australia: 1800 678 242
New Zealand: 0800 507 896

sales@mscience.com.au
www.mscience.com.au

Affymetrix

Revolutionize life

Standardize your drug metabolism studies.

The DMET™ Plus Premier Pack, which includes arrays, reagents and analysis software, is the most comprehensive, cost-effective product containing known pharmacogenetic markers for drug metabolism studies. It standardizes drug metabolism studies and enables faster discovery and measurement of genetic variation associated with drug response compared to traditional methods. The DMET™ Plus Panel features markers in all FDA-validated genes and covers more than 90 percent of the current ADME Core markers as defined by the PharmaADME group.

The DMET Plus Premier Pack is a proven, fast and flexible way to streamline your drug metabolism workflow.

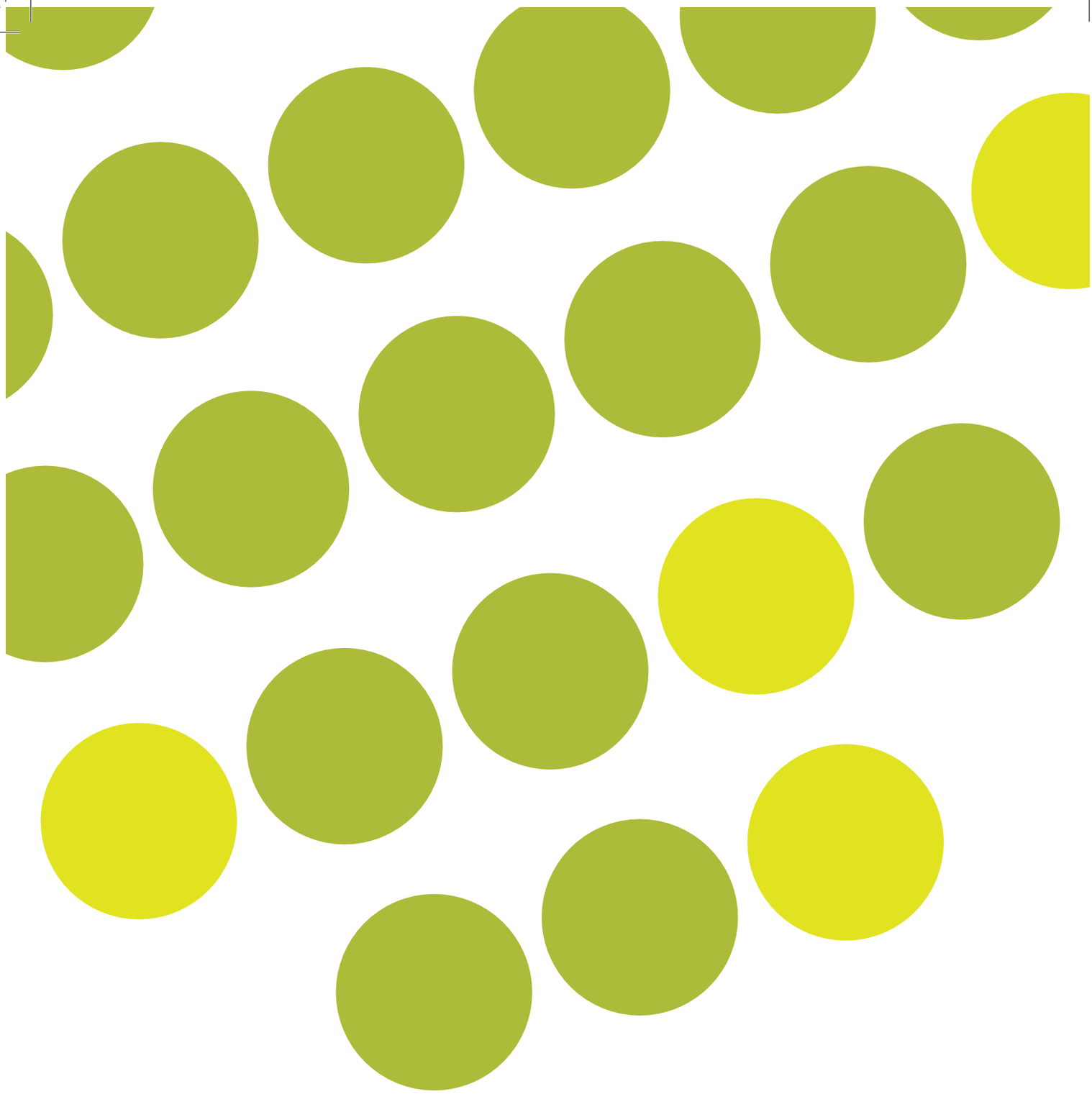
To Learn more come by the Affymetrix/ Millennium Science booth.

millennium science 10 YEARS 1999 - 2009

millennium science

enabling next-generation research

Australia: 1800 678 242
New Zealand: 0800 507 896
sales@mscience.com.au
www.mscience.com.au



KEYNOTE LECTURE

Searching for the dark matter of biological regulation.

Patrick Brown

Department Of Biochemistry, School Of Medicine, Stanford University

What we know about the architecture and mechanisms of biological regulation can account for only a small fraction of the biological phenomena that we can observe. Similarly, only a small fraction of the sequences in the human genome that are under purifying selection (and therefore presumably contain important information) encode proteins or participate in known regulatory processes. Work in my lab and others suggests that a significant fraction of the missing "dark matter" is related to post-transcriptional regulation – including regulation of RNA decay, translation and localization in the cell. Genome-wide studies of protein-RNA interactions, using DNA microarrays and deep sequencing, have uncovered strong evidence that a large and diverse group of RNA binding proteins act combinatorially to control the post-transcriptional fate of every mRNA.

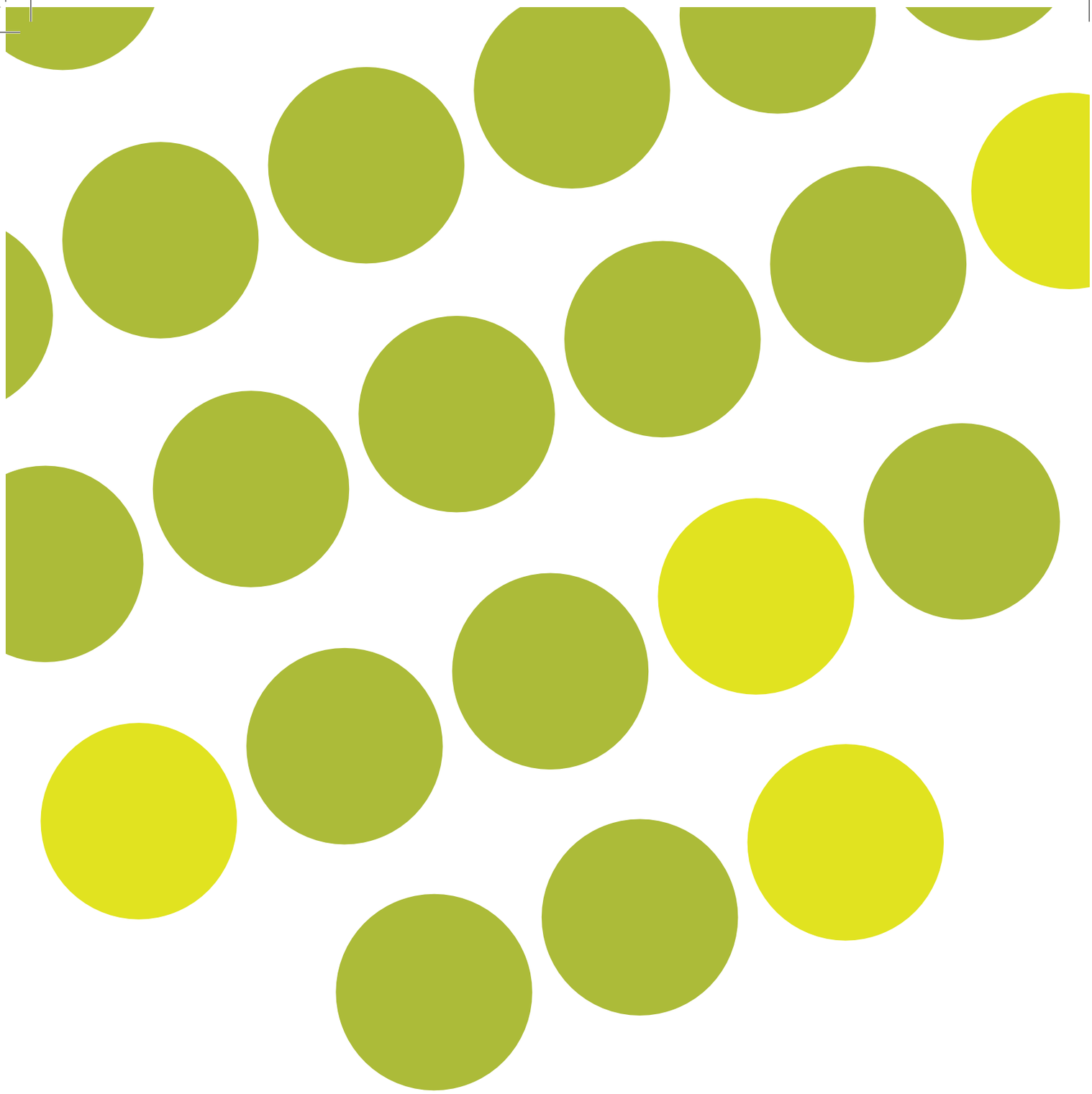
Simplify Adoption of RNA Sequencing with the SOLiD™ System

No matter how you add it up, only the SOLiD™ System delivers the highest sensitivity, greatest dynamic range and the largest catalog of easy-to-use reagent kits. Applied Biosystems—your answer for RNA sequencing.

Learn how to adopt RNA sequencing in your lab by visiting **solid.appliedbiosystems.com/rna**

- Download pioneering publications
- View interactive on-demand webinars
- Compare microarray to sequencing data
- Browse catalog of easy-to-use Ambion® kits

applied biosystems



PLENARY LECTURES

Geographical Genomics, canalization, and the origins of human disease

Greg Gibson

Center for Integrative Genomics, Georgia Institute of Technology, Atlanta

Genome-wide association studies with transcript abundance in peripheral blood samples or derivative cell lines have demonstrated a preponderance of regulatory polymorphisms, also known as eSNPs, which impact the expression of several percent of all genes. Several of these highlight associations that contribute to a variety of disease conditions, but the question arises as to how the associations are affected by the environment. We have addressed the robustness of eSNP associations to decanalization of the transcriptome in the face of different biotic and abiotic challenges faced in different geographic locations. I will describe a gene expression GWAS that controls for population structure and lifestyle, in a comparison of Arab and Amazigh individuals from a city and two villages in southern Morocco. Approximately 400 genomewide significant associations are observed in leukocyte samples obtained from 194 individuals, the vast majority in cis, and all are consistent across the three sample locations and after controlling for ethnicity and relatedness, despite substantial divergence in the structure of the transcriptome in rural villagers. No evidence for large-effect trans-acting mediators of the pervasive environmental influence is found and instead genetic and environmental factors appear to act in a largely additive manner. I will discuss the implications for the origins of complex disease in human societies undergoing profound transitions where genotype-by-environment interactions might be expected to influence disease risk.

Genomics of extinct and endangered species

S.C. Schuster[1], V.M. Hayes[2] and W. Miller[1]

1: Center for Comparative Genomics and Bioinformatics, Penn State University, University Park, USA

2: Cancer Genetics Group, Children's Cancer Institute Australia for Medical Research, Sydney

Only 200 years ago the concept of static species was challenged by the first fledgling theories of evolution. Since then, ever-growing collections of fossils have allowed mankind to gain insight into the constant changes that have shaped fauna and flora for eons. While these studies initially were of strictly anatomical nature, it was discovered only in the last 25 years that in addition to petrified structural information, also biomolecules have survived the demise of individuals and species. With successful sequencing of DNA retrieved from the Quagga, an extinct species of zebra, the field of ancient DNA was invented in 1984 (1). The last three years have seen a rapid succession of improvements in sequencing of ancient DNA , driven by the onset of next-generation sequencing. This has now resulted in a draft version of the mammoth's nuclear genome (2), together with an extensive set of complete mitochondrial sequences of this extinct group of Proboscideans (3). The availability of a large set of genetic information from an extinct species allows for the assessment of the genetic diversity of animals that ceased to exist several 10,000's of years before our time, and thereby allowing the investigation the contributions made by genetic factors to the extinction process. For these analyses, mitochondrial markers have historically been used, as the survival of nuclear DNA on a larger scale had not been documented in fossils until recently. Studies on mammoth populations have revealed a surprisingly small genetic diversity, as well as the existence of two previously undetected groups of animals (3). While this in itself does not explain the extinction process, it raises the question of a contributing factor in addition to the population size. We have investigated this lead in addition to a detailed analysis on the mammoth populations in the recently extinct species of the Tasmanian Tiger (4), who's genetic analysis has been unsuccessfully been attempted for the last decade. Our findings have lead us to believe that the observations made for the two extinct species come to play also in endangered species that are at the brink of extinction today. We are therefore sequencing the nuclear genome of the Tasmanian Devil, the largest remaining marsupial carnivore. This species is currently dramatically endangered through a form of infectious cancer. We will show how the lack of biological diversity in this species is relevant to the animal's failing immune response in the onset of the disease.

The lessons learned from past extinctions, documented through the efforts of today's paleogenomics, may therefore help to assess not only the status of endangerment of a species, but also help it from going extinct by directing breeding programs that are already underway.

Literature

(1) Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC. DNA sequences from the quagga, an extinct member of the horse family. Nature. 1984 Nov 15-21;312(5991):282-4.

(2) Miller W, Drautz DI, Ratan A, Pusey B, Qi J, Lesk AM, Tomsho LP, Packard MD, Zhao F, Sher A, Tikhonov A, Raney B, Patterson N, Lindblad-Toh K, Lander ES, Knight JR, Irzyk GP, Fredrikson KM, Harkins TT, Sheridan S, Pringle T, Schuster SC. Sequencing the nuclear genome of the extinct woolly mammoth. Nature. 2008 Nov 20;456(7220):387-90.

(3) Gilbert MT, Drautz DI, Lesk AM, Ho SY, Qi J, Ratan A, Hsu CH, Sher A, Dalén L, Götherström A, Tomsho LP, Rendulic S, Packard M, Campos PF, Kuznetsova TV, Shidlovskiy F, Tikhonov A, Willerslev E, Iacumin P, Buigues B, Ericson PG, Germonpré M, Kosintsev P, Nikolaev V, Nowak-Kemp M, Knight JR, Irzyk GP, Perbost CS, Fredrikson KM, Harkins TT, Sheridan S, Miller W, Schuster SC. Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. Proc Natl Acad Sci U S A. 2008 Jun 17;105(24):8327-32. Epub 2008 Jun 9

(4) Webb Miller, Daniela I. Drautz, Jan E. Janecka, Arthur M. Lesk, Aakrosh Ratan, Lynn P. Tomsho, Mike Packard, Yeting Zhang, Lindsay R. McClellan, Ji Qi, Fangqing Zhao, M. Thomas P. Gilbert, Love Dalén, Juan Luis Arsuaga, Per G. P. Ericson, Daniel H. Huson, Kristofer M. Helgen, William Murphy, Anders Götherström, Stephan C. Schuster, The mitochondrial genome sequence of the Tasmanian tiger. Genome Research, 19(2), 213-220. Epub 2009 Jan 12.

High-resolution DNA binding specificity profiles of transcription factors and *cis* regulatory codes in DNA

Martha Bulyk

Brigham & Women's Hospital and Harvard Medical School

The interactions between TFs and their DNA binding sites are an integral part of the regulatory networks within cells. Identification of the DNA binding specificities of sequence-specific transcription factors (TFs) is important for understanding transcriptional regulatory networks, in particular for the prediction of cis regulatory modules (i.e., transcriptional enhancers), inference of cis regulatory codes, and interpretation of in vivo TF occupancy data and gene expression data. My group developed highly parallel in vitro microarray technology, termed universal protein binding microarrays (PBMs), for the characterization of the sequence specificities of DNA-protein interactions at high resolution. Using universal PBMs, we have determined the DNA binding specificities of >500 TFs from a wide range of species. These data have permitted us to identify novel TFs and their DNA binding site motifs, predict the target genes and condition-specific regulatory roles of TFs, predict tissue-specific transcriptional enhancers, investigate functional divergence of paralogous TFs within a TF family, investigate the molecular determinants of TF-DNA 'recognition' specificity, and distinguish direct versus indirect TF-DNA interactions in vivo. Further analyses of closely related TFs and the cis regulatory elements in which their binding sites occur are likely to reveal features of cis regulatory codes important for driving appropriate gene expression patterns.

Causes and consequences of microRNA dysregulation in cancer

Muller Fabbri, M.D.

Department of Molecular Virology, Immunology and Medical Genetics Comprehensive Cancer Center

The discovery of microRNAs (miRNAs) has revolutionized the dogma of molecular biology, according to which a gene is first transcribed in a messenger RNA (mRNA), then translated into a protein. miRNAs are small, non-coding RNAs which regulate gene expression by interacting with mRNAs and controlling gene expression both at a transcriptional and post-transcriptional level. Over the past several years it has been demonstrated that miRNA genes contribute to the pathogenesis of most- if not all- human malignancies. Dysregulation of miRNA expression can be driven by various mechanisms, including amplifications, deletions, mutations of miRNA loci, as well as by epigenetic silencing or dysregulation of transcription factors that target specific miRNAs. Since the development of a malignant phenotype is dependent on the dysregulation of miRNA genes, which in turn control or are controlled by the dysregulation of multiple protein-coding oncogenes or tumor suppressor genes, these small RNA molecules provide a rationale for the development of new anti-cancer therapies.

Integrated analysis of the cancer genome

Javed Khan

Center for Cancer Research, National Cancer Institute, Gaithersburg, MD

Neuroblastoma is an extremely heterogeneous disease in which the outcome can range from spontaneous regression of the tumor to relentless progression leading to the death of the patients. For the first half of my talk I will discuss how genomics, including DNA, mRNA and microRNA profiling, has been applied and integrated to decipher the biology of this enigmatic cancer. For the second half of the talk I will discuss the application of next generation sequencing (NGS) techniques, for investigating the cancer genome. NGS technology directly identifies billions of nucleic acid species in parallel in a single experiment. Different from the Sanger method of traditional sequencing, the massively parallel DNA sequencing technology not only generates sequence information for each nucleic acid strand, but also determines the abundance of each nucleic acid species due its large capacity, resulting in a digital readout of levels for any sequence, even those at low levels beyond the detection sensitivity of hybridization-based technologies. I will discuss how this technology has wide-ranging applications for both DNA and RNA studies. For DNA it is possible to sequence an entire cancer genome in one month, a staggeringly short time considering that it took 13 years to sequence a handful of human genomes by the HGP. There are also hybridization based methods for pulling down the DNA from protein coding exons or a defined genomic region (termed 'genome-partitioning') for targeted resequencing. With these methods it is possible to detect every single nucleotide variant, mutation, genomic rearrangement, profile the whole "methylome", and determine copy number alteration at the base pair level of a given genome. For RNA studies it is possible to determine the gene expression level of every gene, identify every splice variants, novel transcripts, single nucleotide variants and mutations for the expressed genome. It will also identify novel gene rearrangements that result in chimeric fusion gene products. Finally I will discuss how next and next-next generation sequencing has and will revolutionize the field of cancer genomics.

Analyzing genome-wide transcription and chromatin with microarrays and next-generation sequencing

Vishy Iyer

Institute for Cellular & Molecular Biology, The University of Texas at Austin

The transcriptional reprogramming of a genome in response to signals involves the coordinated regulation of hundreds of genes, mediated by transcription factors, changes in chromatin structure, and the action of regulatory RNAs. We are using genomic and molecular approaches to understand genome-wide regulatory networks involving transcription factors and their downstream targets, some of which can themselves be regulators. I will describe a gene regulatory network involving the transcription factor Myc and a target miRNA, which in turn promotes cell proliferation by suppressing the interferon response in quiescent cells. In other work being carried out as part of the ENCODE Project, we are analyzing the chromosomal targets of the transcription factor and chromatin organizer CTCF in different cell types. The use of next-generation sequencing (ChIP-seq) also allows us to assess the allele-specific binding of this transcription factor in related human samples, yielding insights into the relation between genetic variation and gene regulation.

Understanding tissue-specific gene expression and disease from large microarray compendia

Olga Troyanskaya

Department of Computer Science & Lewis-Sigler Institute for Integrative Genomics, Princeton University

The ongoing explosion of new technologies in functional genomics offers the promise of understanding gene function, interactions, and regulation at the systems level. However, the complexity and scale of human molecular biology make it difficult to integrate this body of data, understand it from a systems level, and apply it to the study of specific pathways or genetic disorders. These challenges are further exacerbated by the biological complexity of metazoans, including diverse biological processes, individual tissue types and cell lineages, and by the increasingly large scale of data in higher organisms.

I will describe how we address these challenges through the development of bioinformatics frameworks for the study of gene function and regulation in complex biological systems, thereby contributing to understanding of human disease. Specifically, I will describe SPELL, a "Google-like" search engine for very large gene-expression compendia, and HEFalMP, a regularized Bayesian integration system we developed that provides maps of functional activity and interactions in over 200 areas of human cellular biology and disease, each including information from ~30,000 genome-scale experiments pertaining to ~25,000 human genes. These systems allow prediction of protein function and functional modules, cross-talk among biological processes, and association of novel genes and pathways with known genetic disorders. I will also describe our work in starting to model these systems-level processes in a cell-type/tissue specific context, starting with accurate predictions (and experimental confirmation) of tissue-specific expression.

For more information, please explore:

http://function.princeton.edu/

function.princeton.edu/spell

and

function.princeton.edu/hefalmp

Genetic control of growth and survival

David Gresham

Center for Genomics and Systems Biology, Department of Biology, New York University

The regulation of growth programs that are appropriate for particular environmental conditions is essential for the long-term survival of free-living microbes. In order to meet this requirement, networks of interacting genes must assess the external environment and establish the appropriate transcriptional, translational and metabolic states in the cell. I will discuss two approaches to interrogating these networks. First, I will describe studies of adaptive evolution under nutrient-limited growth conditions using chemostat cultures. In these long-term selections, mutants that have improved growth out-compete the wildtype ancestral strain. In order to identify the complete set of mutations in evolved mutants, I developed a method of genome-wide mutation detection using tiling DNA microarrays. I have now extended this method to the reliable detection of heterozygous mutations through the use of microarrays with isothermal, variable probe length microarrays. Second, I will discuss studies in which I have used reverse genetics to simultaneously measure the survival of all ~4,500 haploid yeast deletion mutants when starved for different nutrients. I have developed a new method, using quantitative Solexa sequencing, to accurately assess mutant abundance on the basis of unique molecular barcodes. This method, CODE-SEQ, provides a powerful and accurate means for quantitative phenotyping of heterogeneous mutant pools enabling genetic screens of phenotypes that were previously impractical for large-scale analysis.

GUEST LECTURES

Single cell quantitative gene expression analysis from zygote to blastocyst

G. Guo[1,3], E. M. Huss[2], N. Clarke[2], P. Robson[1,3]

[1]Stem Cell & Developmental Biology and [2]Computational & Systems Biology, Genome Institute of Singapore, Singapore 138672
[3]Department of Biological Sciences, National University of Singapore, Singapore 117543

The first cellular differentiation event in mouse development leads to the formation of the blastocyst consisting of the inner cell mass (ICM) and a functional epithelium (trophectoderm; TE). The ICM shortly thereafter gives rise to the pluripotent epiblast (EPI) and the extra-embryonic primitive endoderm (PE). The molecular mechanisms that regulate the differentiation of totipotent blastomeres to the three distinct cell types remain unclear. Here we apply quantitative single cell technology to profile the expression of 48 genes, in parallel, from 576 individual cells harvested throughout preimplantation development. We found that blastomeres at the 16-cell stage abundantly express numerous transcription factors that subsequently become lineage-restricted. However, zygotic Sox2 is specific to the 16-cell stage inner cells; while Id2 is specific to the outer ones. In the early blastocyst (~32 cells) TE cells are readily distinguishable from ICM, and by the ~64-cell stage, EPI from PE. An inverse correlation between Fgf4 and Fgfr2 expression in early stage ICM cells suggests differential expression of these paracrine signaling components is an early event in the development of the EPI and PE. These results dramatize the power of quantitative single cell expression analysis to provide insight into developmental mechanisms and should be widely applicable to other biological systems.

Keywords: Single cell gene expression profiling, Mouse blastocyst, Preimplantation development.

ChIA-PET, a whole-genome approach to study human chromatin interactome in transcription regulation

Xiaoan Ruan

Genome Institute of Singapore, 60 Biopolis Street, Singapore 138672

The human genome is organized into high level structures, and DNA elements separated by large genomic distances could functionally interact. Many transcription factors regulate transcription by binding to remote regulatory DNA elements distant from gene promoters. Therefore, the study of protein-DNA interactions and the long-range interactions between regulatory sites, collectively called the chromatin interactome, would illuminate important aspects of genome biology in healthy and disease cells. However, chromatin looping and their impact on transcription regulation are not fully understood and have not been investigated in a genome-wide manner. Current methods for studying chromatin interactions including 3C are limited to a few interactions per experiment. To understand remote control of gene transcription in the context of entire genome, we developed a strategy for Chromatin Interaction Analysis using Paired End diTag sequencing, called ChIA-PET. In this approach, chromatin interactions are captured by formaldehyde cross-linking, followed by sonication to fragment chromatin fibers and to separate the non-specific chromatin fragments from specific interaction complexes that are further enriched by chromatin immunoprecipitation. Tethered DNA fragments in each of the chromatin complexes are connected with DNA linkers through proximity ligation. The ligation products are subjected for paired-end-tag extraction, which are analyzed by ultra-high-throughput sequencing. The resulting ChIA-PET sequences are mapped to the reference genome to reveal the relationships between remote chromosomal regions brought together in spatial proximity by protein factors. We have done extensive validation to prove that the ChIA-PET data is reliable, accurate, and reproducible for de novo identification of chromatin looping.

Using this high-throughput and genome-wide ChIA-PET approach, we have comprehensively mapped the chromatin interaction network mediated by estrogen receptor α (ERα) and other protein factors in a human cancer genome. We found that the great majority of remote ERα binding sites are anchored to the promoter regions of target genes through long-range chromatin interactions. Our data suggests that ERα functions in transcription regulation by bringing genes together through extensive looping structure of chromatin interactions into transcriptional foci, which could facilitate the recruitment of other specific or general transcription factors for transcription initiation and the maintenance of local high concentration of transcription components for active transcription. Our findings suggest a likely possibility that chromatin interaction is a primary mechanism for regulating transcription in mammalian genomes, particularly in transcriptional induction.

Ppie: a modifier of epigenetic reprogramming in the mouse identified by NimbleGen sequence capture and 454 sequencing.

Nadia Whitelaw[1], Nicola Vickarous[1], Jeffrey Jeddeloh[2], Emma Whitelaw[1]

1: Queensland Institute of Medical Research

2: Roche NimbleGen, USA

In an effort to identify and study epigenetic modifier proteins in the mouse, we have performed a sensitised ENU mutagenesis screen. Using an epigenetically sensitive variegating GFP transgene as a reporter, we have uncovered a dozen dominant mutant lines that exhibit altered transgene variegation. These are named Mommes (modifiers of a murine metastable epiallele). The causative point mutation has been identified in seven lines. Mutations have been found within the DNA methyltransferase Dmnt1 and Snf2h, a chromatin remodeller, which validate the screen. Mutations have also been found in relatively novel genes, for example SmchD1 and Baz1b.

In all mutant lines homozygous lethality is observed, in some cases with incomplete penetrance. In one line, MommeD8, some homozygotes die after birth. Viable homozygotes weigh less and have increased DNA methylation at the GFP transgene. MommeD8 was mapped to a 4 Mb interval on chromosome 4. Traditional candidate gene sequencing of exons failed to find the causative mutation and so Roche NimbleGen Sequence Capture technology and 454 Sequencing was used to sequence the entire MommeD8 interval. Homozygous and heterozygous resequencing data was compared to a de novo wildtype reference assembly. Two point mutations were identified in mutants, one was intergenic and the other within an intron of the Peptidylprolyl isomerase E (Ppie) gene. MommeD8 mutants display reduced levels of full length Ppie mRNA making this the likely causative mutation. Our results suggest that Ppie plays a role in epigenetic gene silencing.

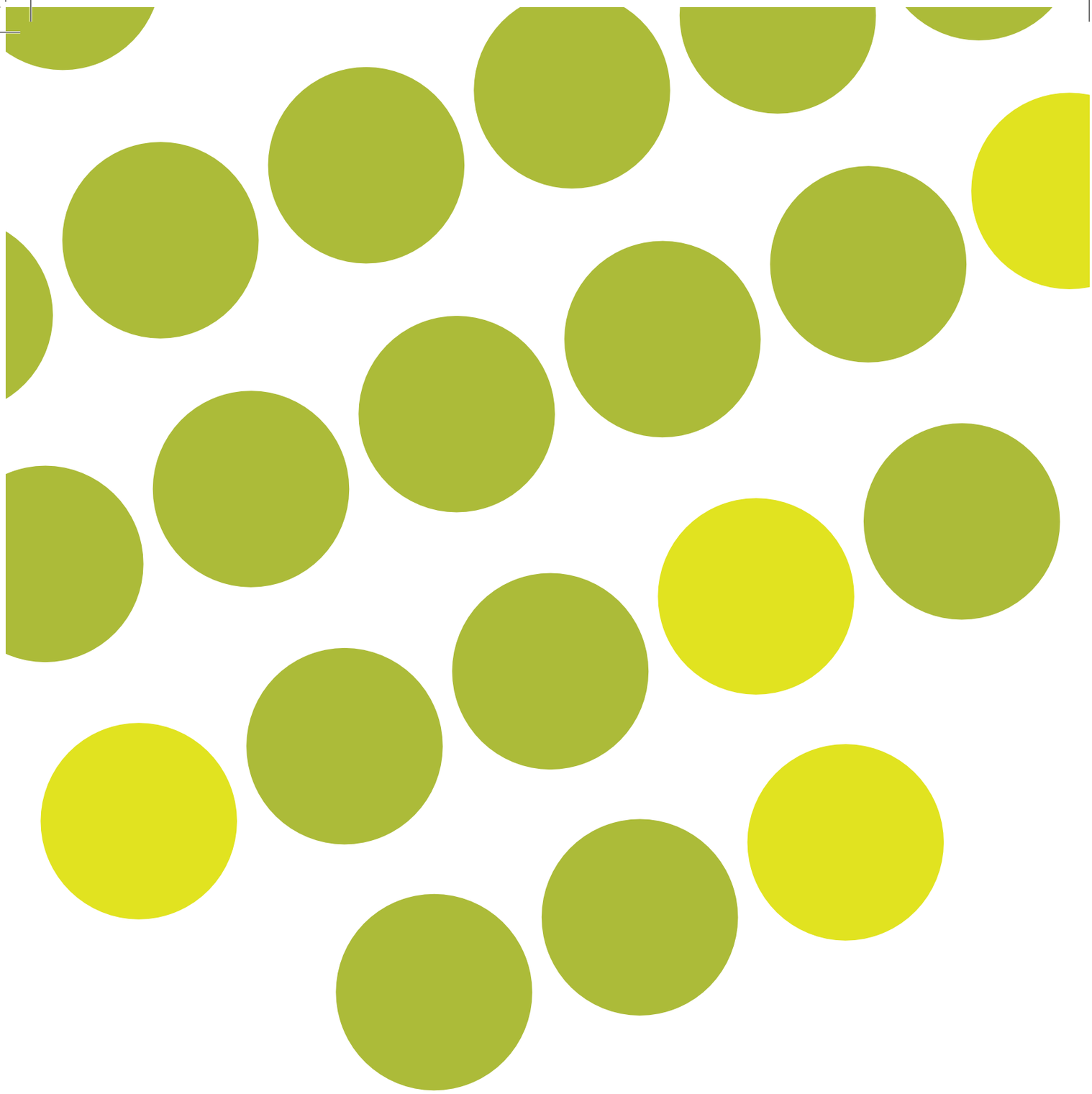
Submitted by Nadia Whitelaw - nadia.whitelaw@qimr.edu.au

Integrated analysis of genomic regulatory regions using cell arrays and next-generation sequencing.

Michal Janitz

School of Biochemistry & Biomolecular Sciences, University of NSW, Sydney, NSW, Australia

The promoter regions integrate information about the status of the cell in which they reside and alter the rate of transcriptional initiation of a single gene accordingly. It has been realized that mammalian promoters are too diverse to allow reliable computational annotation of genomic DNA sequence without reference to experimentally determined locations of full-length cDNA sequences or transcription start sites (TSSs). Furthermore, identified promoter regions need to be functionally evaluated in order to deliver a comprehensive picture of their regulatory role. To that end, a comprehensive functional analysis of over 180 promoters derived from human chromosome 21 (Chr21) will be presented. Promoter activity information, generated using transfected-cell arrays, was compared with that of the gene expression data measured by quantitative RT-PCR and digital gene expression (Illumina Genome Analyzer-based DGE). In addition, systematic bioinformatics analysis was performed for all the putative promoters. The integrated mapping of active promoter elements for the HSA21 genes and the associated gene regulatory elements will be discussed.



SYMPOSIA

Combining next generation high throughput sequencing and population genomics

Carsten Külheim, Suat Hui Yeoh, Gavin Moran, and William Foley

Research School of Biology, Australian National University

Australian Myrtaceae are well known for their high content of foliar essential oils, which are mostly a mixture of terpenoids. There is huge interest in understanding the quantitative and qualitative variation in these compounds within and between plant species. This interest comes partly from the industry, who farm certain Myrtaceae species, such as Melaleuca alternifolia and Eucalyptus polybractea for their essential oils and partly from basic scientific interest for the understanding of plant – herbivore interactions. Herbivores such as Brushtail possums (***Trichosurus vulpecula***), Ringtail possums (***Pseudocheirus peregrinus***) or Koalas (***Phascolarctos cinereus***) choose their food source depending on the composition of certain terpenoids and terpenoid-adducts.

We have characterized the oil composition and quantities of a comprehensive collection of Eucalyptus globulus trees. We can now associate allelic variations with the oil phenotype. We have used two different approaches: 1. Re-sequencing of genes of interest from pooled DNA with a next generation sequencer, followed by SNP genotyping with the Sequenom MassARRAY and 2. Direct re-sequencing of genes of interest, amplified from individuals and barcoded prior to sequencing on a next generation sequencer. The latter approach holds many advantages; it is much cheaper, faster and provides better resolution of allelic variants. When using the MassARRAY, only selected single nucleotide polymorphisms can be assayed, whereas with the second method, all variants within the genes of interest can be identified and analyzed simultaneously.

Submitted by carsten kulheim - carsten.kulheim@anu.edu.au

Discovery of single nucleotide polymorphisms in sugarcane for gene mapping using 454 sequencing

Bundock PC, Eliott FG, Ablett G, Benson AD, Bowen S, and Henry RJ

CRC for Sugar Industry Innovation through Biotechnology, Centre for Plant Conservation Genetics, Southern Cross University

Sugarcane is genetically complex due to polyploidy, aneuploidy and hybridisation which have led to a large but variable number of copies of each chromosome and a large number of chromosomes overall in the genomes of commercial varieties. To map genes of interest, 'single dose' SNPs are desirable because they are fully informative across the mapping population. However discovery of these SNPs requires considerable depth of sequencing to find the single copy alleles and confirm the polymorphisms. To achieve this depth, 454 sequencing of pooled PCR amplicons was utilised on the parents of a QTL mapping population. Three hundred pooled amplicons from each parent were sequenced yielding 96,755 and 86,241 sequences from the two parents, with average sequence depth of approximately 300 and average read length of 220 bases. In the more polymorphic parent, 94% of amplicons analysed (227/242) had evidence of a reliable SNP – an average of a SNP every 35 bases. Candidate single dose SNPs were validated and genotyped for mapping across the progeny using the Sequenom MassARRAY (MALDI-TOF mass spectrometer) system. From 225 candidate SNP sites tested, 209 (93%) were validated as polymorphic using the Sequenom system. Genotyping across the mapping population was carried out for 197 SNPs. Amplicon re-sequencing using the 454 system enables cost effective SNP discovery that can be targeted to genes of interest. This approach should be useful for the detection of SNPs in polyploid species generally - for linkage mapping, association studies and for population & ecological genetics.

Submitted by Peter Bundock - peter.bundock@scu.edu.au

Measuring inter-individual variation in gene expression in monoamine neurotransmitter pathway genes

Vicky Cho[1,2], Simon Easteal[2,3] and Rohan Williams[1,2],

1: Molecular Systems Biology Group, ANU

2: Genome Biology Program, John Curtin School of Medical Research, Canberra, ACT, John Curtin School of Medical Research, Canberra, ACT

3: Predictive Medicine Group, ANU

Understanding the origin of differences in global gene expression between individuals, whether arising from either genetic and/or environmental sources, is a problem of current interest. Regardless of whether genetic and/or environmental factors are identified as influencing gene expression in a given setting, the patterns of gene expression related to inter-individual differences still need to be represented and understood, or in other words, how can we best capture or represent how individuals in a study population differ in the patterning of their gene expression changes? Here, we develop the concept of inter-individual co-variance to capture this level of variation in gene expression,. Rather than attempt to analyse global expression differences between individuals, we focus on analysing inter-individual differences that are present in gene-sets, (i.e. a set of genes that share some common property e.g. co-regulation, membership of a known signalling pathway, etc). We illustrate inter-individual expression variation in genes in dopamine and serotonin related pathways using expression data from temporal cortex of 135 human subjects. Whilst a similar range of inter-individual variability was present amongst genes in the two pathways, many genes showed high levels of inter-individual variability, for example the schizophrenia candidate gene RGS4 showed a 22-fold change in expression (relative to its mean) across the set of 135 subjects. We further illustrate how principal components analysis of these data can provide direct visualisation of expression variation in gene sets. These approaches will be applicable to understanding the patterning of expression changes resulting from genetic and/or environmental perturbations.

Submitted by vicky cho - vicky.cho@anu.edu.au

Massively parallel sequencing of epigenomic and functional genomic landscapes

Izhak Haviv[1,2,3] , Wen Qiu[2] , Ian Campbell[2,3] , Adam Kowalczyk(3,4), Richard Tothill[2] , Terry Speed[5] , Min Hu[6] , Kornelia Polyak[6] Alex Dobrovic[2] , Stephen Fox[2] , Assam El-Osta[1] , Bryan Beresford-Smith[4] , Thomas Conway[4]

1: Baker IDI

2: Peter MacCallum Cancer Centre

3: University of Melbourne

4: National ICT Agency

5: Walter and Elisa Hall Institute

6: Dana Farber School of Medicine, Harvard University

We used GAII to interrogate multiple repeats of a number of biological comparisons in tumour progression (DCIS versus normal), diabetes (endothelial cell response to glucose), and epithelial polarity (RNAi screen of RAS suppressor Scrible). In each of those studies, novel methods of feeding the resulting nucleic acid to the GAII as well as analysis tools were developed. In particular, we describe CpG methylation profiling of archival material following laser capture microdissection, chromatin immunoprecipitation of acetylated-K9 of H3 histone, and target sequencing of Sure-select-enriched 3 megabase fragment of the human genome. Technical trouble shooting and limitations are described.

Submitted by Izhak Haviv - izhak.haviv@bakeridi.edu.au

The transcriptome of an adult coral and its symbiotic zooxanthellae characterized using 454 pyrosequencing

Souter, P.[1], Vera, JC.[2], and van Oppen, MJH[1].

1: Australian Institute of Marine Science, PMB no 3, Townsville MC, Queensland 4810, Australia

2: Department of Biology, 208 Mueller Laboratory, Pennsylvania State University, University Park, PA 16082, USA

Genomic scale studies of non-model organisms that lack fully sequenced genomes have recently become a possibility. As a consequence, the scientific community is starting to develop molecular markers that are targeted to address ecological issues such as resilience, adaptation and fitness. This study presents the result from the sequencing and de novo assembly of the adult coral Acropora millepora together with its algal symbiont, Symbiodinium sp. Extracted mRNA from eight individuals from each of two thermally distinct populations were pooled for the analysis to improve coverage and increase the chance of detecting polymorphisms that are unique to a single population and therefore worth targeting as markers for thermal stress tolerance. A total of 215,747 sequences were generated in this initial run. These assembled into 2,438 contigs, leaving 14,997 singletons. Average depth of coverage ranged from 1.2 – 27.73 and 3.2 for the longer contigs (>300 bp). Four-hundred-and-fifty-five of the contigs generated strong BLAST alignments to Uniprot protein database, 388 of which were from unique gene annotations. Additionally, 40 of the contig alignments were found to be from the algal symbiont, of which 18 were unique. An initial SNP screening found 336 candidates. A second 454 titanium sequencing run is currently being conducted which will add to these results.

Submitted by Petra Souter - p.souter@aims.gov.au

Identifying cause of Cerebellar Abiotrophy in Australian Kelpie dogs using SNP arrays, Nimblegen sequence capture arrays and 454 sequencing.

Jeremy R. Shearman and Alan N. Wilton

School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW 2052, Australia and Clive and Vera Ramaciotti Centre for Gene Function Analysis, University of New South Wales, Sydney, NSW 2052, Australia

Dogs are mankind's longest running breeding experiment resulting in huge physical and behavioural variation between breeds. Modern dogs have undergone several major population bottlenecks, domestication from the wolf and individual breed creation. These bottlenecks have resulted in limited genetic variation and large haplotype blocks, ten times the length of human blocks allowing mapping to be performed with fewer markers than human. This makes dogs ideal for the genetic study of diseases and physical/behavioural traits.

We used the Canine Affymetrix SNP array v2 for a whole genome analysis (WGA) to identify the location of the gene for Cerebellar Abiotrophy (CA) which is a common autosomal recessive disorder that causes ataxia in the Kelpie, 14 affecteds and 23 controls were genotyped. Homozygosity analysis identified a 5 Mb region which is identical in all affecteds but is also a common haplotype in the population. Out of the 44 genes in this region the most likely to be involved in ataxia showed no sequence variation. NimbleGen sequence capture arrays for the entire 5 Mb region were used to enrich genomic DNA from 2 affecteds and 1 control dog. Quantitative PCR confirmed enrichment of each sample. The 3 samples will be tagged using multiplex identifiers and pooled for a 454 Titanium sequence run. Sequence analysis will be performed using GS De Novo Assembler and GS Reference Mapper. Once the mutation has been identified, a diagnostic test will be designed and made available for breeders to identify carriers of the disorder.

Widespread alternate transcription in human monocyte derived macrophages

Anthony G. Beckhouse[1], Alistair M. Chalk[1], Yu-lan S. Lo[1], Carole Juruj[1], Jiyuan An[1], Katharine M. Irvine[2], Matthew J.Anderson[1], Christine A. Wells[1],

1: Eskitis Institute for Cell and Molecular Therapies, Griffith University, Nathan, Australia

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

The innate immune system is the primary defence barrier against systemic infections. When a pathogen is detected, a rapid response is elicited by components of the innate immune system to remove the threat. The macrophage is an important cell within this system that detects pathogens through germline encoded pattern recognition receptors such as the Toll-like receptors (TLRs). Each receptor recognises a unique pathogen pattern that activates a highly conserved signalling cascade and results in a specific inflammatory response.

Alternate splicing activity in mouse TLR signalling has been identified previously, therefore we hypothesised that the same process may be essential to the diversification of inflammatory signalling in man. The current study utilised Affymetrix All Exon arrays to identify alternate transcription events in human macrophages challenged with lipopolysaccharide (LPS). Macrophages were differentiated ex vivo from peripheral blood mononuclear cells originating from 4 unrelated healthy donors. RNA was isolated from cells over a time-course of treatment with LPS. Array data was analysed using the FIRMA software package and identified many genes with inducible expression of one or more cassette exons.

It was predicted that of the 70 core members of the TLR signalling cascade, 3 protein variants were expressed from each gene. The predicted variant ORFs are undergoing systematic qPCR validation, cloning, tagging and over-expression in CHO and COS7 cells to determine localisation and phenotypic changes associated with presence of the variant. We conclude that expression of alternate transcription mediates the diverse signalling response through key TLR signalling components.

Submitted by Anthony Beckhouse - a.beckhouse@griffith.edu.au

Beyond the exome: Re-interpretation of genome-wide association studies using an empirically-based annotation of the noncoding transcriptome

Marcel E. Dinger, Tim R. Mercer, and John S. Mattick

Institute for Molecular Bioscience, University of Queensland, St Lucia 4072.

Although protein-coding sequences account for less than 2% of the human genome, recent studies show the genome to be pervasively transcribed mostly as long noncoding RNAs (lncRNAs). Initially considered transcriptional 'noise', an increasingly wide range of functions are being ascribed to lncRNAs including the ability to regulate gene expression by acting as transcriptional co-activators, targeting epigenetic modifications, and mediating post-transcriptional regulation. These roles underscore their contribution to disease aetiology and development, and we have shown the expression of specific lncRNAs to be associated with a wide range of disease states and developmental processes, including melanoma, breast cancer, HIV infection, embryonic stem cell differentiation, T-cell activation and neurogenesis. The recognition that lncRNAs contribute towards disease motivated the comprehensive survey of lncRNA annotation and expression from RNAseq datasets in a range of human tissues. Deep sequencing of the transcriptome (RNAseq) has considerable potential for identifying and determining the expression of lncRNAs, since, unlike probe-based approaches, it provides an unbiased description of the transcriptome. However, RNAseq does not currently identify the full-length of the transcript, a requirement for current bioinformatic methods to discriminate coding from noncoding transcripts. Furthermore, such bioinformatic approaches are limited in their ability to identify exotic or small protein-coding regions, or conversely misannotate open reading frames that occur in lncRNAs by chance alone. To address these limitations, we have developed a novel empirically-based approach to annotate the noncoding transcriptome. The annotation derives from mapping of ~10 million proteomic tags to the human genome. Coupling peptide sequence data with genomic and spliced EST data can further infer the full-length open reading frame and expand the annotated proteome. This method generates a remarkably expansive proteomic annotation of the human genome, which itself identifies many small, novel and exotic proteins. This approach eclipses all previous methodologies to discriminate coding and noncoding transcripts since it does not require a priori knowledge of full-length transcripts, and does not suffer the limitations of theoretical-based approaches that can only provide statistical likelihoods for the existence of putative open reading frames. This approach can be incorporated to annotate coding and noncoding regions of any RNAseq dataset regardless of depth. We have applied our approach to provide the most accurate annotation of the human transcriptome to date and reveal the specific expression of large numbers of novel lncRNAs in both normal and disease human tissue. In addition to intergenic lncRNAs, lncRNAs are also often found neighboring protein-coding genes of clinical importance, organised in a bidirectional, intronic, antisense or overlapping manner, each with their own implications to genesis and regulation. The global identification of lncRNAs prompts a reconsideration of disease-linked SNPs in noncoding regions of the genome, a subset of which we show occur within lncRNAs. The accurate distinction between coding and noncoding expression in the genome is critical for the interpretation of disease-risk SNPs and development of therapies and prognostic and diagnostic assays. Our annotation comprises an initial step in considering the majority of disease risk-variants that lie beyond the exome and will add considerable insight and value to personal genome sequencing and genome-wide association studies.

Submitted by Marcel Dinger - m.dinger@uq.edu.au

Sequencing the transcriptional complexity of small RNAs

Nicole Cloonan[1,2,3], Qinying Xu[1,2,3], Ehsan Nourbkhsh[1,2,3], Keerthana Krishnan[2,3], Brooke Gardiner[1,2,3], Gabriel Kolle[2,3], John Mattick[2,3], Sean Grimmond[1,2,3]

1: Queensland Centre for Medical Genomics

2: Institute for Molecular Bioscience

3: The University of Queensland

The small-RNA transcriptome (<35nt) can mediate a diverse range of gene-regulatory functions. MicroRNAs (miRNAs) are capable of interfering with target gene activity. Endogenous small interfering RNAs (siRNAs) and Piwi-associated RNAs (piRNAs) are involved in retrotransposon silencing, and small nucleolar RNAs (snoRNAs) mediate chemical modification events on RNA. High-throughput sequencing approaches have revolutionized our ability to survey the content and complexity of small-RNA populations, and one of the most striking observations reported is the detection of multiple miRNA species generated from the same precursor hairpin.

We have exhaustively sequenced small-RNA fractions from 10 human tissues to obtain a thorough estimate of the content, complexity and spatial dynamics of the human small-RNA transcriptome. A rigorous identification and classification system to define isomiRs on the basis of length, variant 5' seed sequence, and overall sequence content was developed, and we observe far more complexity from known miRNAs than previously reported. Many isomiRs were both tissue restricted and independent of canonical miRNA abundance. These data form the foundation for addressing the scale, scope, and complexity of small-RNA expression, and suggest that we are far from understanding the true level of transcriptional complexity in the small RNA population.

Submitted by Nicole Cloonan - n.cloonan@uq.edu.au

The role of basic Krüppel-like factor (Klf3) in B-lymphocyte development and tumorigenesis

Richard C.M. Pearson, Thi Thanh Vu, and Merlin Crossley

School of Molecular and Microbial Biosciences, University of Sydney

Basic Krüppel-like factor (Bklf), also known as Klf3, is a transcriptional repressor that is highly expressed in haematopoietic cells. Bklf is implicated in B-cell cancer. Most notably retrovirus integrations upstream of the Bklf gene are associated with B-cell lymphomas in mice.

We have been studying the Bklf knockout mouse and have identified several defects in B-cell development. In the Bklf null bone marrow, the percentage of pre/pro-B-cells is increased with a similar reduction in early pre-B-cells. We observe an approximate 3-fold increase in the percentage of mature cells, accompanied by a significant reduction in the percentage of immature late pre-B cells. This suggests that Bklf may play a critical role during the pro- to pre-B-cell transition, as well as in the later B-cell maturation process, perhaps by influencing both pre-B cell receptor and B-cell receptor signalling pathways. On exiting the bone marrow, B-cells migrate to the spleen, where important cell fate decisions determine whether B-cells differentiate into either follicular or marginal zone cells. In the absence of Bklf, we observe a 60-70% reduction in the number of marginal zone cells.

We are now seeking to identify Bklf target genes in specific B-cell populations at key stages of development, using Affymetrix microarray technology. The arrays indicate that Klf3 may regulate both proliferative and apoptotic networks in the developing B cell. It is hoped that understanding these transcriptional networks will give insight into normal and malignant B-cell growth.

Submitted by Richard Pearson - richard.pearson@usyd.edu.au

Bioinformatics of high throughout RNA data.

William Ritchie, Justin Wong, and John Rasko

Centenary Institute

High throughput genomic technologies produce server size amounts of data. To stock, analyze and interrogate these data, Bioinformaticians use mathematical and computational methods borrowed from a variety of fields such as economy, sociology or linguistics. Because the data produced from biological experiments is unique in size and quality, Bioinformatics is emerging as an independent discipline with unique approaches. In my talk I will discuss why these approaches are unique to Bioinformatics and will illustrate through two examples how these may advance our knowledge of biological systems. 1) Scaleability: Here I demonstrate how Bioinformatics approaches gain from zooming out from the specific details of a biological mechanism to larger-scale analysis and how Bioinformaticians need to expertly switch between both levels of detail. 2) Transversality: Here I demonstrate why Bioinformatics requires a precise knowledge of techniques and approaches from multiple disciplines. I will show through the example of an in-house webtool how this knowledge applies to a seemingly simple task: visualizing microRNA expression profiles from on-line data. I will finish my talk by answering the question: Do you need Bioinformatics in your lab? This open question is designed to focus on the differences between simple statistics feasible on an excel spreadsheet, biostatistics and Bioinformatics.

Submitted by william ritchie - W.Ritchie@centenary.org.au

Systematic discovery of gene regulation networks from genome-wide transcription factor binding motif data

Oscar Junhong Luo[1], Xiaohui Xie[2], and Rohan Williams[1]

1: The John Curtin School of Medical Research, Australian National Univeristy

2: Department of Computer Science, University of California, Irvine

Understanding the organization, regulation and function of high eukaryotic gene expression programs in development, homeostasis and disease remains a major goal of system biology research. Collections of defined transcription factor (TF) binding motifs and their corresponding target genes in a given species can be used as a "dictionary" to define groups of genes that have the potential of being under co-ordinated control. Using a collection of genome-wide motif presence data in the human genome from Xie et al., 2008, comprising 440 distinct TF binding motifs and their 17,238 target genes, and, by using the concept of "transcriptional control similarity" and hierarchical clustering, identify groups of genes plausibly under co-ordinated transcriptional control. We subsequently focused on the 1,352 "stable" clusters that were reproducible under a resampling based cluster validity test. 88% of these stable clusters share more than 1 motif (range: 2 to 53, median: 5) that are annotated to all genes in the cluster. Furthermore, in only less than 10% of these stable clusters, all genes were located on the same chromosome. Approximately half of the stable clusters showed ubiquitous expression patterns across 73 non-cancerous tissues (SymAtlas data). We performed Gene Ontology Biological Process terms enrichment analysis to examine the hypothesis that co-regulatory potential can imply co-functionality. Three clusters were enriched (Bonferroni corrected) for the GO terms: regulation of transcription (DNA-dependent), macromolecule metabolic process and developmental process, respectively. These findings demonstrate the feasibility of identifying putative trans-regulated clusters of genes using pre-defined motif dictionaries and simple multivariate approaches.

Submitted by Oscar Junhong Luo - Oscar.Luo@anu.edu.au

Small RNA profiling of the HL-1 cardiomyocyte cell line using next generation sequencing

David Humphreys[1], Arthur Liu[1], Grace Wei[1], Romaric Bouveret[1], Richard Harvey[1,2], Thomas Preiss[1,2]

1: Victor Chang Cardiac Research Institute

2: University of New South Wales

MicroRNA (miRNA) are 19 – 24 nucleotide non-coding RNA that can repress translation of target mRNAs by interaction with partially mismatched sequences in their 3'UTR. Because miRNA function is dependent on sequence, knowing the detailed cellular transcriptome composition of a biological sample is valuable for interpreting the function of miRNA. Next Generation Sequencing (NGS) technology can provide researchers with this level of information. We have used SOLiD NGS to obtain a detailed sequence profile (>25million reads) of small RNA (<35nt) from the mouse HL-1 cardiac cell line. We profiled RNA sampled over a four-day time course, where cells were plated and started to grow in a non-beating state, through to confluency, where they beat in synchrony. This time course may mimic the progression of a non-differentiated cardiac tissue (ie non-beating) through to a differentiated one (beating). We found that the HL-1 cell line highly expresses many miRNAs known to be important in heart development (eg miR-1, -133, -29a, -30), but also other, previously unknown small RNAs. Our bioinformatic analyses of these data are still ongoing but have already revealed much information on alternative processing and other sequence variations of miRNAs (e.g. processing into the 'star' form, occurrence of isomiRs, and editing) and other small RNA found to be expressed in the HL1 cell line. Our current focus is on utilising our dataset to identify small RNAs that are differentially expressed during the differentiation time course. Ultimately, we wish to complement this small RNA profiling data set with NGS data on mRNA expression to discover novel miRNA-target interactions that are important to cardiac biology.

Submitted by David Humphreys - d.humphreys@victorchang.edu.au

Identification of miRNA regulatory modules and their targets using matched miRNA-mRNA data

Vivek Jayaswal[1,2] and **Jean Yee Hwa Yang**[1,2]

1: School of Mathematics and Statistics, University of Sydney

2: Sydney Bioinformatics, Centre for Mathematical Biology University of Sydney

MicroRNAs (miRNAs) are short non-coding RNA molecules that act as post-transcriptional regulators of mRNA. Since miRNAs can act as oncogenes or tumour suppressors, an understanding of their regulatory mechanism could provide new insights into cancer and other diseases. One major challenge in the identification of target-mRNAs is the need to accommodate the many-to-many mapping between miRNAs and mRNAs – a miRNA can target hundreds of mRNAs and several miRNAs target a single mRNA. We propose a statistical approach for determining the miRNA regulatory modules and their target-mRNA modules. This approach is applicable to matched miRNA-mRNA data, with the matched data being either a timecourse data or a data spanning multiple tissues.

For mRNA data, we use multivariate random forest to obtain an N×N similarity matrix, where N denotes the number of mRNAs. The similarity matrix is determined based on the mRNA expression values and the computationally predicted regulators (miRNAs) of mRNA. We perform a soft thresholding of the large similarity matrix followed by tight clustering to obtain the mRNA modules. The same process is repeated for miRNA data to obtain miRNA regulatory modules. Finally, we measure the association between miRNA and mRNA modules and determine clusters of miRNAs that target clusters of mRNAs. A simulation study is performed to determine whether the majority of mRNAs in a given module are indeed targeted by a small set of miRNAs. The preliminary results are encouraging even for low accuracy of predicted miRNA targets.

Submitted by Jean Yang - jeany@maths.usyd.edu.au

A large scale RNAi screen for regulators of scribble function in cancer reveals the mammalian cell polarity network as a novel regulator of Ras oncogenic function.

Lorey K Smith[1,5*], Lukas E Dow[1*?], Nathan J Godde[1*], Allison J Ogden[1], Tanja Schlosser[1], Ryan Galea[1,6], Jason Li[3], Imogen Elsum[1,6], Helen Pearson[1], Raj Gaire[7], Gad Abraham[7], Kathy Jastrzebski[4], Ross Dickins[9], Carleen Cullinane[2], Sherene Loi[2,10,11], Izhak Haviv[7,8], Patrick O Humbert[1,5,6]

1: Cell Cycle and Cancer Genetics Laboratory

2: Translational Research Laboratory

3: Bioinformatics Core

4: Protein Chemistry Laboratory, Peter MacCallum Cancer Center, East Melbourne, VIC 3002, Australia

5: Department of Pathology

6: Department of Molecular Biology and Biochemistry

7: NICTA, Department of Computer Science and Software Engineering

University of Melbourne, Parkville, VIC 3052, Australia

8: Baker IDI Heart and Diabetes Institute, Melbourne, VIC 3004, Australia

9: Walter and Eliza Hall Institute, Parkville, VIC 3052, Australia

10: Department of Translational Research and Functional Genomics, Jules Bordet Institute, Boulevard de Waterloo 121, Brussels 1000, Belgium

11: Department of Medicine, St Vincent's Hospital, University of Melbourne, Victoria, Australia

Scribble is a core polarity regulator and tumour suppressor. In Drosophila, loss of Scribble cooperates with activated oncogenes such as Ras or Notch to give rise to invasive and metastatic tumours. We have now shown that the invasive and tumour suppressive properties of the mammalian homologue of Scribble are highly conserved. Loss of human Scribble is able to cooperate with activated H-Ras to drive invasion of mammary epithelial cells and conversely overexpression of human Scribble can suppress Ras-mediated cell transformation phenotypes including loss of apico-basal polarity, invasion in 3D culture and importantly, anchorage independent growth. How Scribble regulates oncogenic Ras signaling is currently unknown.

Here we describe an unbiased large-scale, functional genomics screen to identify the genetic requirements for the tumour suppressive functions of Scribble. Specifically, we have queried which genes are required for the ability of human Scribble to suppress Ras-driven cell transformation, combining RNAi technology with Next Generation Sequencing. We have screened ~20,000 shRNA constructs from the Open Biosystems genome-wide miR30-based lentiviral shRNA library, maintained at the Victorian Centre for Functional Genomics. Through this approach, we have identified 141 candidate genes that are potentially important regulators and effectors of Scribble's tumour suppressive functions. Subsequent primary and secondary validation experiments on a number of key targets confirmed not only their requirement for Scribble's inhibition of Ras-driven oncogenesis, but also their ability to cooperate with oncogenic Ras to drive highly invasive phenotypes in vitro. Importantly, the gene target set was highly enriched for the majority of currently known core polarity regulator complexes including those involved in asymmetric cell division and the control of spindle pole localization. Of note, the mammalian homologue of the Pins asymmetric cell division regulator, GPSM2/LGN was one of the key targets of the screen and we provide evidence for the tumour suppressive function of GPSM2 in vivo in mice and in the progression of human breast cancer.

Submitted by Lorey Smith - lorey.smith@petermac.org

The influence of epigenetic state on genetic variation

Hua Ying[1], Julien Epps[2], Rohan Williams[1], **Gavin Huttley**[1]

1: JCSMR, ANU

2: School of Electrical Engineering and Telecommunications, UNSW

Multi-cellular organisms rely on epigenetic modifications to control the timing and level of gene transcription. Epigenetic modifications exert a functional influence by interfering with DNA-binding of regulatory factors, either directly in a localised manner or more broadly by affecting the degree of DNA compaction. Experimental studies establish that chromatin compaction affects rates of both DNA lesion formation and repair. A functional association between chromatin status and 5-methyl-cytosine also exists. These suggest that both the total rate and type of substitution will be affected by chromatin status. Regular positioning of nucleosomes, the building block of chromatin, further predicts that substitution rate and type should vary spatially in an oscillating manner. We addressed chromatin's influence on substitution rate and type in primates. Matched numbers of sites were sampled from Dnase I Hypersensitive (DHS) and closed chromatin control Flank. Likelihood ratio tests revealed significant excesses of total and of transition substitutions in Flank compared with matched DHS for both intergenic and intronic samples. An additional excess of CpG transitions was evident for the intergenic, but not intronic, regions. Fluctuation in substitution rate along ~1800 primate promoters were measured using phylogenetic footprinting. Significant positive correlations were evident between the substitution rate and a nucleosome score from human T-cells, with up to ~50% of the variance in substitution rate accounted for. Using signal processing techniques, a dominant oscillation at ~200 bp was evident in both the substitution rate and nucleosome score. Our results suggest that epigenetic modifications impose a substantial mutation cost on eukaryotic organisms.

Submitted by Gavin Huttley - Gavin.Huttley@anu.edu.au

Comparative global gene expression patterns of colonic epithelium in azoxymethane-induced carcinogenesis in rats and colorectal cancer in humans.

Caroline A. Kerr[1,2], Janet Shaw[1,2], Robert Dunne[1,3], Lauren M. Bragg [1,3], William J. Wilson[1,3], Julie Clarke[1,4], Trevor Lockett[1,2], Richard Head1.

1CSIRO Preventative Health Flagship CSIRO, 2CSIRO Food and Nutritional Sciences, North Ryde, NSW, 3CSIRO Mathematical and Information Sciences, North Ryde, NSW and 4CSIRO Food and Nutritional Sciences, Adelaide, SA, Australia.

Azoxymethane (AOM) induced carcinogenesis in rats is an important model of human colorectal cancer (CRC). Using microarrays, gene expression changes in the rat colonic epithelium immediately after AOM injection (Acute model) and in AOM-induced tumours (Cancer model, 27 weeks after exposure) were compared to those occurring in human colorectal cancer. In rats, global gene expression in normal proximal versus distal rat colonic tissue differed markedly in both models. Rat tumours from different regions of the colon showed limited positional variability in gene expression. Therefore, loss of proximal/distal 'site of origin' signatures in tumours is an important consequence of AOM-induced carcinogenesis. In the Acute model, AOM exposure induced substantial changes in gene expression in the colonic epithelium. Whilst some of these changes were still apparent in the Cancer model, they were considerably reduced; suggesting repair has occurred in the damaged epithelial cells that did not progress into tumours. The tumour gene expression was radically different to the apparently 'normal' colonic epithelium. A comparative genomics in silico tool was developed to compare the rat microarray data derived from normal and carcinogen exposed colonic epithelium or tumour tissue with related data from positionally-matched human CRC and normal colonic epithelium. Gene expression networks were devised for both models. There was some overlap (41% and 21% respectively) when differentially expressed genes in the rat Acute and Cancer models were compared to their human orthologs differentially expressed in cancer. Networks prepared for these conserved genes further highlight similarities between these rat models and human CRC.

Transcription factor GATA2 in the repression of adipogenesis

Briony H.A. Jack and Merlin Crossley

University of Sydney

The development of mature adipocytes from preadipocyte precursor cells requires co-ordinated changes in gene expression. Genes responsible for cellular proliferation are downregulated and genes that contribute to lipid biosynthesis and storage are upregulated. Management of these expression changes relies on the actions of DNA-binding transcription factors 1.

Gata2 has been identified as a negative regulator of the early stages of adipogenesis. Gata2 is expressed in preadipocytes and its expression is down regulated at the onset of adipogenesis. Constitutive expression of Gata2 in the 3T3-L1 preadipocyte cell line blocks their differentiation.

We have investigated the role of Gata2 and its corepressor partner proteins, Friend of Gata and C-terminal Binding Protein (CtBP) in suppressing adipogenesis. Our results indicate that these corepressors play important roles in regulating adipogenesis. CtBP is of particular interest as it responds to NADH and functions as a metabolic sensor. Our results therefore provide a framework whereby the entry into the adipocyte programme may be regulated by the metabolic status of preadipocytes. We are currently using microarray technologies to investigate the particular target genes regulated by Gata2 and its partners.

1. Sue, N, Jack, B, et.al. (2008) Mol. Cell. Biol. 28(12),3967-78.

Submitted by Briony Jack - b.jack@usyd.edu.au

RNA-seq enables whole genome allele specific expression

Matthew J Wakefield and Alicia Oshlack

Walter and Eliza Hall Institute, Melborne

RNA-seq measures of transcription by fragmentation of mRNA and sequencing on high throughput sequencing platforms. RNA-seq provides an expression measure for each transcript and opportunity for the discovery of novel transcripts, alternative splicing, and measuring allele specific expression at transcribed SNPs. To achieve the promise of RNA-seq an understanding of the data and new statistical methods are critical (Oshlack & Wakefield 2009 Biology Direct 4:14).

To validate, optimize, and determine the limits of detection of allele specific expression on the Illumina GAII we have conducted a known truth mixture experiment using tissues from two strains of mice. Our mixture design allows calculation of the true expression level and ratio of alleles for individual samples free from unknown biological effects. We will present results examining the effects of sequence mapping, statistical analysis and sequencing depth on the recovery of the truth using allele specific RNA-seq data.

We show that RNA-seq can be used as an accurate and sensitive platform for discovery of allele specific expression with applications in epigenetics, cancer, eQTLs and genomic conflict in hybrids.

Submitted by Matthew Wakefield - wakefield@wehi.edu.au

Differential regulation of tissue specific gene expression by modular assembly of enhancers

John E. Pimanda[1], Kathy Knezevic[1], Mary E. Janes[1], Rabab Nasrallah[1], Aileen Smith[2] , Sarah Kinston[2], Anja Kolb-Kokocinski[3], Katrin Ottersbach[2], Georges Lacaud[4], Valerie Kouskoff[4] and Berthold Göttgens[2].

1: Lowy Cancer Research Centre and the Prince of Wales Clinical School, University of New South Wales, Sydney, Australia.

2: Department of Haematology, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK.

3: The Wellcome Trust Sanger Institute, Cambridge, UK.

4: Paterson Institute for Cancer Research, Christie Hospital, Manchester, UK.

Endoglin is an accessory receptor for TGF-b signalling and is required for normal hemangioblast, early hematopoietic and vascular development. We have previously shown that an upstream enhancer, Eng-8 together with the promoter region mediates robust endothelial expression yet is inactive in blood. To identify hematopoietic regulatory elements, we employed array based methods to determine chromatin accessibility across the entire locus. Subsequent transgenic analysis of candidate elements showed that an endothelial enhancer at Eng+9 when combined with an element at Eng+7 functions as a strong hemato-endothelial enhancer. ChIP-chip analysis demonstrated specific binding of Ets factors to the promoter as well as to the -8, +7 and +9 enhancers in both blood and endothelial cells. By contrast Pu.1, an Ets factor specific to the blood lineage, and Gata2 binding was only detected in blood. Gata2 was bound only at +7 and GATA motifs were required for hematopoietic activity. To test whether these elements are active in the bi-potential haemangioblast and to establish the binding motifs that are responsible for their activity, wild-type and mutant enhancers have been targeted to the HPRT locus of HM1 ES cells for blast colony forming cell assays. We believe that the modular assembly of regulators give blood and endothelial progenitors the regulatory freedom to independently fine-tune gene expression during development and emphasizes the role of regulatory divergence in driving functional divergence.

Data mining the genetics of leukemia

Geoff Morton[1], David Skillicorn[1], and Daniel R. Catchpoole[2]

1: School of Computing, Queens University, Kingston Ontario, Canada

2: The Tumour Bank, The Oncology Research Unit, The Children's Hospital at Westmead, NSW, Australia

Acute Lymphoblastic Leukemia (ALL) is the most common cancer in children under the age of 15. Presently, all diagnosis, prognosis and treatment decisions are made based upon blood and bone marrow laboratory testing. With advancements in microarray technology it is becoming more feasible to perform genetic assessment of individual subjects as well. We used Singular Value Decomposition (SVD) on merged Illumina SNP with Affymetrix and cDNA gene expression data and performed aggressive attribute selection using random forests to reduce the number of attributes to a manageable size. We then explored clustering and prediction of patient specific properties such as disease sub-classification, and especially clinical outcome. We determined that integrating multiple types of data can provide more meaningful information than individual datasets if combined properly. We have identified the most useful integration method of for combining these datasets together which is to first combine the datasets and then perform attribute selection using random forests. This method is able to capture the correlation between attributes. The most striking result is an apparent connection between genetic background and patient mortality under existing treatment regimes. We find that we can cluster very well based upon the mortality label of the patients. Also, using a Support Vector Machine (SVM) we can predict the clinical outcome for each patient in our study cohort with 100% accuracy. This presentation will discuss the data mining methods used and their application to biomedical research, as well as our results and how this will affect the diagnosis, prognosis and treatment of ALL in the future.

ChIP-seq chromatin modification data facilitates accurate prediction of tissue-specific transcription factor binding sites

Tom Whitington, Andrew C. Perkins, and Timothy L. Bailey

Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072, Australia

Despite the critical role of transcription factors (TFs) in determining the transcriptional output of a cell, current computational models are not able to accurately predict TF binding locations in eukaryotic genomes. In order to understand and predict TF binding, computational methods must be developed to incorporate the salient parameters, such as chromatin structure and co-factor activity.

Chromatin immunoprecipitation combined with high-throughput sequencing and tiling microarrays has recently enabled genome-wide mapping of chromatin modifications and transcription factor binding sites (TFBSs) in multiple organisms and tissues. These datasets facilitate for the first time a genome-wide characterisation of the relationship between chromatin modifications and TF binding. To this end, we have investigated the utility of high throughput chromatin datasets in computational TFBS prediction, for seven human and mouse tissues, using 28 human and mouse ChIP-seq and ChIP-chip datasets as a gold standard to evaluate predictive performance.

We show that incorporating high-throughput chromatin modification estimates greatly improves the accuracy of computational prediction of in vivo binding for a wide range of transcription factors in human and mouse. This improvement is superior to the improvement gained by equivalent use of either transcription start site proximity or phylogenetic conservation information. Importantly, predictions made with the use of chromatin structure information are tissue-specific.

This result supports the biological hypothesis that chromatin modulates transcription factor binding to produce tissue-specific binding profiles in higher eukaryotes, and suggests that the use of chromatin modification information can lead to accurate tissue-specific transcriptional regulatory network elucidation.

Submitted by Tom Whitington - t.whitington@imb.uq.edu.au

Investigating the erythroid role of the transcription factor Klf3

Alister Funnell, Richard Pearson, Ka Sin Mak, and Merlin Crossley

University of Sydney

Krüppel-like factors (Klfs) are transcription factors that bind to CACCC-boxes and related sequences in the control regions of their target genes. The 17 mammalian Klfs regulate the expression of distinct, and in some cases overlapping, sets of target genes in a wide range of biological processes. The founding member of the Klf family, Klf1, is predominantly expressed in erythroid cells and has been shown to drive the expression of genes that are required for normal red blood cell development, such as adult β-globin. We have shown that the expression of another family member, Klf3, is highly dependent on Klf1 in erythroid cells. Moreover, we have characterised the Klf3 locus and identified an alternative promoter which is mainly active in erythroid cells. Taken together, these results suggest that Klf3 may play a role in erythropoiesis. We have subsequently utilised a Klf3 knockout mouse model in order to explore this possibility. Klf3 null mice, while viable, are mildly anaemic. They have increased nucleated red blood cells, both in utero and perinatally, and an elevated number of reticulocytes (immature red blood cells) and Howell-Jolly bodies (red blood cell inclusions). Microarray analysis has revealed a cohort of genes that are deregulated in Klf3 null Ter119+ (erythroid) E14.5 foetal liver cells. The majority of these genes are upregulated, consistent with the view that Klf3 acts primarily as a transcriptional repressor. Together with the phenotypic observations, the microarray data are forming a picture of perturbed erythropoiesis in the Klf3 knockout mice.

Submitted by Alister Funnell - a.funnell@usyd.edu.au

Genuine association of gene expression profiles

Belinda Phipson and Gordon K. Smyth

The Walter and Eliza Hall Institute

Consider a typical microarray experiment in functional genomics in which two or more genes have been perturbed (by mutation, knockout or other means) and are compared back to a common wild type. Our interest is to infer whether the perturbed genes affect common molecular pathways. An analysis may involve determining the number of significant genes in each mutant genotype compared to the wild type. A common approach to relate independent profiles is to look at Venn diagrams and assign p-values using contingency table tests, but this is not appropriate when the comparisons are computed from the same data. In this case the expression profiles are not independent and are in fact correlated due to the comparison to the common wild type.

Instead of looking at counts of genes, the log fold changes become the focus; which avoids an arbitrary cut-off. If the perturbed genes are affecting the same genes and pathways, the log fold changes should change together. However if there are no common genes or pathways, the log fold changes should change independently.

An extension to the hierarchical empirical Bayes model attempts to separate true biological correlation from the technical correlation observed when comparing to common wild type groups. This method allows the inclusion and estimation of a correlation coefficient for pairs of comparisons. It has been applied to various datasets, in particular one that involves mutations affecting platelet counts, and another that affects PRC complexes.

Submitted by Belinda Phipson - phipson@wehi.edu.au

Building transcriptional regulatory networks for heart development

Bouveret R., Doan T., and Harvey R.P.

The Victor Chang Cardiac Research Institute

In humans, structural and functional malformations of the heart are very common and are associated with a high economic and emotional burden. In our lab, we are taking a Systems Biology approach to study how genetic networks initiate and control heart development at a molecular level.

Despite considerable advance in the understanding of molecular mechanisms that control development of the mammalian heart, it remains essential to identify most if not all the direct target genes of key transcription factors that play a major role in this process. For this purpose, we have adapted the DNA adenine methyltransferase identification (DamID) method. This technique is a fairly new method developed as an alternative to chromatin immuno-precipitation to facilitate the identification in vivo of DNA loci that interact with specific nuclear proteins both in cell cultures and in whole organisms. We combined the DamID technique with Affymetrix promoter and whole-genome microarrays for the genome-wide identification of new target genes. Using this method, we have been able to: 1) identify and compare target genes of more than eight key cardiac transcription factors; 2) analyse the distribution of transcription factor binding sites both at the gene and genome levels; and 3) study the effect of missense mutations or deletions on the DNA-binding activity and specificity at the genome level.

In conclusion, we are using DamID not only to build gene regulatory networks that control heart morphogenesis but also to help better understand transcriptional genetic networks important for development.

Submitted by Romaric Bouveret - r.bouveret@victorchang.edu.au

Modeling the molecular cell system of hepatocellular carcinoma as an actor-semiotic network

David C.Y. Fung and Marc R. Wilkins

1: New South Wales Systems Biology Initiative, School of Biotechnology and Biomolecular Sciences, The University of New South Wales, New South Wales, Australia

Primary hepatocellular carcinoma (HCC) is currently the fifth most common malignancy and the third most common cause of cancer mortality worldwide. Because of its high prevalence in developing nations, there have been numerous efforts made in the molecular characterization of primary HCC. However, a better understanding into the pathology of HCC requires software-assisted network modelling and analysis. We present an actor-semiotic network model for studying the biological implication of gene co-expression in HCC. Our model is a bipartite graph containing binary node sets. The actor nodes represent physical entities, i.e. gene, protein, microRNA and GO_Component terms. The semiotic nodes represent ontological entities, i.e. GO_Process, GO_Function, Morbid, and Quiescence program. The edges represent either inter-actor relationships or actor-semiotic relationships. The inter-actor relationships are protein-protein, microRNA-gene, protein-GO_Component interactions, and stochastic gene co-expression relationships. The actor-semiotic relationships represented are gene-GO_Process, gene-GO_Function, gene-Morbid, and gene-Quiescence_program.

The resulting network successfully integrates data sources from transcriptomics, proteomics, phenomics, and Gene Ontology. It also captures the molecular cell system of HCC as both a state machine and a complex interactome.

Because the topology of an actor-semiotic network is determined by the combination inter-actor and actor-semiotic relationships, there should be visually identifiable topological features that discriminate between normal hepatocyte and HCC phenotypes. To achieve this, we employed visual graph exploration. In this presentation, we will discuss some important observations we made out of the model. These include co-expression subnetworks, topological positions of microRNA in relation to the co-expression and protein-protein interaction subnetworks, quiescent genes affected, and other diseases that could be related to HCC.

Submitted by David Fung - dcyfung@unsw.edu.au

Development of high-throughput, high-content, functional screening: ARVEC+

Thomas J Gonda[1], Benjamin Wilson[1], Max Ranall[1], Duka Skalamera[1], Amy Purdon[1], Ping Zhang[1], Simon Barry[3], Sean Grimmond[2], Paul Leo[1] and Brian Gabrielli[1]

1Diamantina Institute and 2Institute for Molecular Bioscience, University of Queensland, Brisbane, QLD; 3Department of Paediatrics, University of Adelaide, Adelaide, SA.

Advances in genomics and genetics have allowed the identification of candidate genes associated with biomedically-important processes at a staggering rate. However our ability to attribute functions to such genes, and thus determine their true relevance to the process under study, has lagged behind. This problem, and the potential value of identifying genes based on their functions alone, has spurred the development of a new field and set of technologies collectively termed "functional genomics".

We have initiated a functional genomics program under the acronym "ARVEC" which aims to establish high-throughput cell-based gain-of-function screening. Its initial and still major aims are to 1 generate an arrayed lentiviral cDNA expression library representing the majority of human protein-coding genes and 2 establish a high-throughput platform for carrying out and analysing, using high-content imaging, phenotype-based screens of this library. The scope of the program has recently been expanded to include RNA interference-mediated loss-of-function screening.

This presentation will outline the project and the approaches being taken to achieve these aims. Progress to date will be discussed, including our first screens of a ~1,500 gene pilot cDNA library for cell cycle S phase-promoting genes. Important issues identified during these screens, and some of the challenges of data analysis will be discussed. Finally some intriguing preliminary results from these screens will be presented.

Characterisation of intra-specific genomic diversity in industrial microorganisms by whole-genome next-generation sequencing.

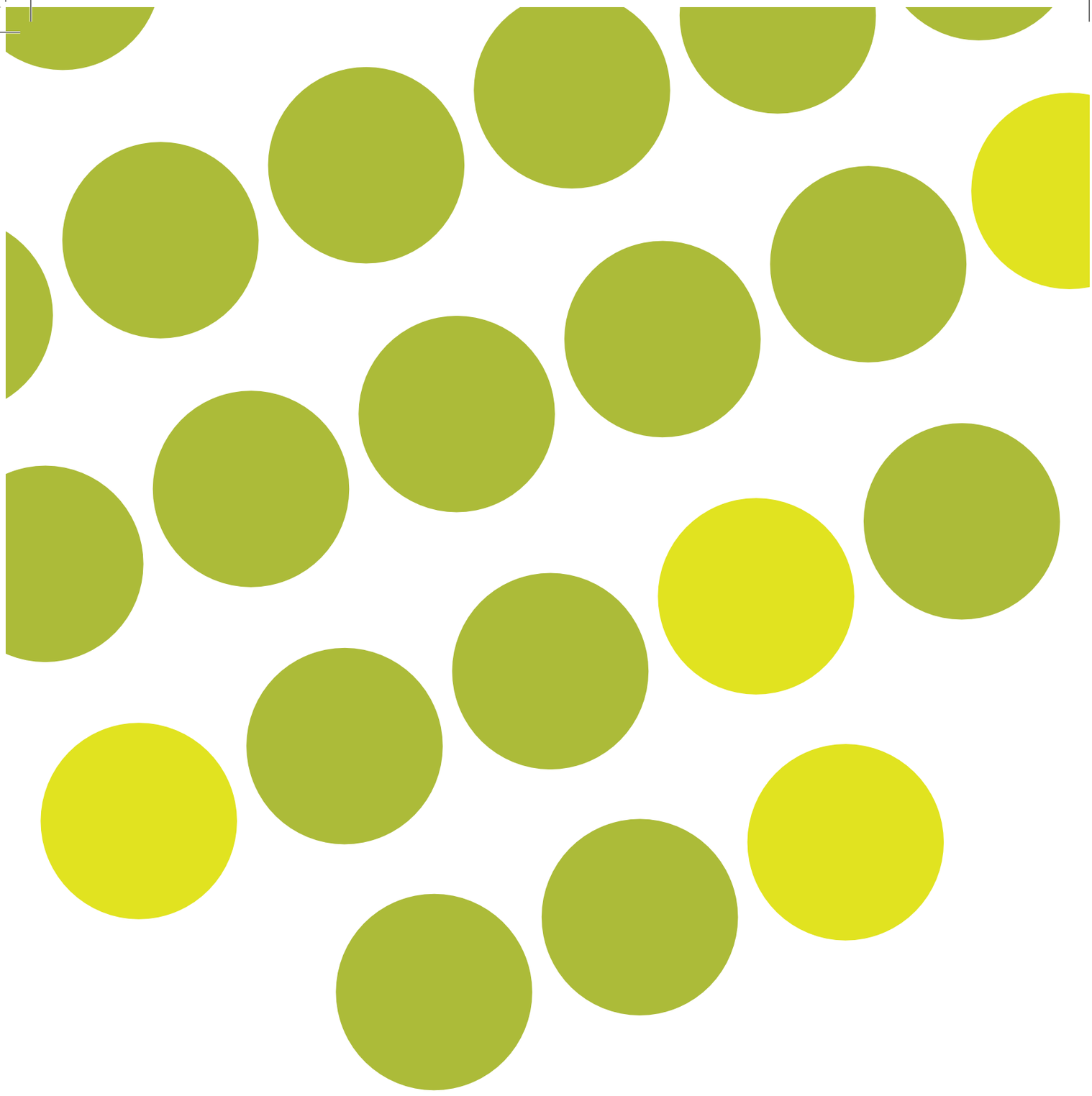
Anthony R Borneman[1], Eveline J Bartowsky[1], Jason P Affortit[2], Isak S Pretorius[1], Michael Egholm[2] and Paul J Chambers[1]

1: The Australian Wine Research Instiute

2: 454 Life Sciences

Industrial microorganisms, such as the yeast Saccharomyces cerevisiae that is used in the winemaking, brewing and pharmaceutical industries, are a diverse collection of microorganisms that have been selected for their ability to perform specific biochemical transformations despite exposure to osmotic, nutrient and ethanol stress. In many cases, the phenotype of industrial microbes is highly variable across strains of the same species and by understanding the genetic basis of these phenotypic differences, it will be possible to maximise desirable characteristics within a strain while minimizing undesirable characteristics. We have used next-generation sequencing and comparative genomics to catalogue the variation present across laboratory and industrial strains of the same species in both the yeast S. cerevisiae and the wine bacterium Oenococcus oeni. In each case, genomic data have uncovered a significant pool of genetic diversity within each species. Individual strains were shown to contain large amounts of nucleotide variation (SNPs), while also displaying significant differences in gene content due to the presence of large deletions and strain-specific insertions of novel genes. This genomic data will be combined with transcriptomic, proteomic and metabolomic information in order to associate phenotypic diversity with specific genomic variation and to allow for predictions to be made regarding how the introduction of genomic variation will impact upon specific industrial traits in these and other important industrial species.

Submitted by Anthony Borneman - anthony.borneman@awri.com.au



TRADE SESSION

Affymetrix' next generation automated genotyping platform

Martin J. Goldberg

Senior Vice President, Research & Development

AFFYMETRIX, INC.

DNA microarray technology continues to play an important role in identifying the genetic basis of common and complex human diseases through genome wide association studies (GWAS). As large scale sequencing efforts across multiple populations continue to add to the characterization of both common and rare genetic variation in the form of single nucleotide polymorphisms (SNP) and copy number variants (CNV), microarrays are well poised to capitalize on this new information content. To this end, Affymetrix has continued to develop new technological approaches to facilitate GWAS. Here we will present an overview of our next generation genotyping platform; an integrated, high throughput system that includes a highly automated workflow, a new enzymatic-based assay, and novel SNP content. We will also describe a large diversity screen that was performed and share some of the outcomes of those activities in the context of genomic coverage and array product configurations. In conclusion, Affymetrix' next generation platform for GWAS provides comprehensive and accurate genotyping of hundreds of thousands of SNPs in a single assay and offers a sample throughput coupled with minimal manual intervention consistent with the needs of large scale studies being conducted as part of the search for the underlying genetic basis of complex human disease.

POSTERS

1

Transcriptional Framework and Expression Viewer: A database for integrating multiple gene expression platforms with transcriptional annotations

Matthew J. Anderson[1] , Amanda L. Miotto[1], Nick A. Matigian[1], Alan Mackay-Sim[1], Alistair M. Chalk[1], Christine A. Wells[1]

1: National Centre for Adult Stem Cell Research, The Eskitis Institute for Cell and Molecular Therapies, Griffith University, Nathan, 4111

To obtain a complete picture of gene expression in Adult Stem-cell populations we need to integrate multiple measurements of gene activity from many different platforms. Classic 3' microarrays measure transcript expression, all-exon arrays measure exon usage and alternative splicing, Cap Analysis Gene Expression (CAGE) surveys transcriptional start sites using sequencing based approaches and Single-Nucleotide Polymorphism (SNP) chips identify genomic variation.

To be able to manage, store, visualise and query all relevant Adult stem-cell transcriptome data, we developed a database schema for a transcriptional framework and our expression datasets and constructed a webpage to present the data. The Transcriptional framework describes annotations for probes, genes, transcripts and exons, gene symbols, KEGG pathways and in house gene lists; with mappings of probes to genes, transcripts, exons and pathways; genes to transcripts, exons and pathways. The database includes data from Illumina, Affymetrix, CAGE, and includes measurments of both expression and genotypic variation. As well as alternative splicing predictions based on FIRMA and COSIE.

The website allows users to search for a gene or set of genes by probes, symbols, Entrez genes ids, Ensembl IDs, KEGG pathways and gene lists from the transcriptional framework and retrieve detailed expression measurements and comparisons of populations at a gene, transcript and exon level, for multiple platforms and direct links to external resources.

The website provides a platform for collaboration between stem-cell and computational biologist, and allows in-depth analysis of expression for functionally related sets of genes.

Submitted by Matthew Anderson - m.anderson@griffith.edu.au

2

Gene expression profiling in malignant melanoma

Carla F. Kairupan[1], Nikola A. Bowden[1], Katie A. Ashton[1], Xu Zhang[2], Peter Hersey[2], Rodney J. Scott[1,3]

1: Centre for Information-Based Medicine, University of Newcastle; Discipline of Medical Genetics, School of Biomedical Sciences, Faculty of Health, University of Newcastle and Hunter Medical Research Institute, Newcastle, NSW, Australia

2: Oncology and Immunology Unit, Calvary Mater Newcastle Hospital, Newcastle, NSW, Australia

3: Hunter Area Pathology Service, Newcastle, NSW, Australia

Melanoma is the fourth most common cancer in Australia. The incidence of melanoma is increasing at an alarming rate with the most effective form of treatment being surgical resection of early stages of disease. Other treatment strategies still remain relatively ineffective for metastatic disease, therefore, new methods for early identification and better treatment strategies are warranted. Understanding the genetic basis of melanoma is important for the development of techniques to identify individuals at risk, to improve patient survival rates and to develop personalised treatments. Given the limited knowledge of the genetic mechanisms involved in melanoma, the aim of this study was to identify new molecular mechanisms associated with disease.

Whole genome gene expression analysis (Illumina Ref8 V2) was used to determine gene expression differences between 93 malignant melanomas and 8 control melanocyte cell lines. Gene transcripts with significantly altered expression ($p<0.05$) and greater than 2-fold change were identified for the melanomas compared to the melanocytes (GeneSpring GX V10).

Analysis of the expression profiles with the significantly altered genes revealed that melanomas have a unique gene expression pattern that could be distinguished from the melanocytes. The melanomas displayed feature sets of gene transcripts altered in cell cycle control, DNA repair and apoptosis.

This analysis revealed specific gene transcripts altered in melanoma, allowing for further insight into the biological mechanisms involved in the genesis of this malignancy. These genes can be further interrogated to help in the identification of individuals at risk and to improve treatment options for patients with melanoma.

Submitted by Katie Ashton - katie.ashton@newcastle.edu.au

3

Assigning roles to sequence motifs using comparative genomics

Timothy L. Bailey, Fabian Buske, Denis C. Bauer, and Mikael Boden

Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072, Australia

High-throughput gene expression and genomics data is increasingly enabling the discovery of DNA sequence motifs that are involved in gene regulation. We showed previously that these motifs can be used to predict the functional role of the entities (e.g., transcription factors (TFs) or micro-RNAs) that bind to them. Our approach is to first predict the target genes of the motif by scanning promoter regions, and then perform a statistical analysis for over-represented Gene Ontology (GO) terms associated with the (putative) target genes. This approach is implemented in GOMO, which is part of the MEME Suite of motif-based sequence analysis software. We have now extended GOMO to include comparative genomes in the scan for target genes. This results in an increase of more than 40% in the number of significant GO terms detected per TF motif in human. We have also incorporated a novel method of estimating the false discovery rate (FDR) that relies on shuffling the GO "map" rather than the promoter regions or the motif. This results in a large increase in the speed of the analysis, allowing GOMO to be provided as a free web service. Compared with using a single species, both the sensitivity and accuracy of predicted roles for TF motifs in E. coli, yeast and human are increased using our new, alignment-free comparative genomics approach. Using 56 human TF motifs, at an FDR of 5%, GOMO now predicts an average of more than 50 significant GO terms per TF.

Submitted by Timothy Bailey - t.bailey@imb.uq.edu.au

4

Genome wide gene expression of induced sputum in non-eosinophilic asthma

Katherine J Baines[1,2], Jodie L Simpson[1,2], Rodney J Scott[3], Lisa G Wood[1,2] and Peter G Gibson[1,2]

1: Priority Research Centre for Asthma and Respiratory Disease, School of Medicine & Public Health, University of Newcastle, NSW

2: Department of Respiratory & Sleep Medicine, HMRI, John Hunter Hospital, Newcastle, NSW

3: Priority Research Centre for Information Based Medicine, School of Biomedical Sciences, University of Newcastle, NSW

Rationale:

Induced sputum is an important non-invasive tool used to assess airway inflammation in asthma. Analysis of sputum inflammatory cell counts has lead to the identification and characterisation of non-eosinophilic asthma phenotypes. Non-eosinophilic asthma accounts for up to half of all asthma cases, however the molecular mechanisms remain unknown.

Objective:

To investigate differential gene expression of induced sputum samples collected from subjects with neutrophilic and paucigranulocytic asthma.

Methods:

Induced sputum samples were collected from subjects with non-eosinophilic asthma. Non-eosinophilic asthma was further characterised into neutrophilic (n=12) or paucigranulocytic asthma (n=18) based on the presence or absence of sputum neutrophils (>63%) respectively. Genome wide gene expression profiles were generated from sputum RNA samples using Illumina Sentrix Humanref-8 expression microarrays, and data was analysed using GeneSpring 10. Differentially expressed genes were defined by both significance ($p<0.05$) using unpaired t test and change of greater than 2 fold.

Results:

There were 196 genes differentially expressed in neutrophilic compared to paucigranulocytic asthma, including 186 genes that were upregulated and 10 genes that were downregulated. Genes upregulated in neutrophilic asthma were involved in the innate immune response such as cytokines (***IL1B, OSM***), receptors (***IL8RA, IL8RB***), proteases (***MMP9, MMP25***), signalling molecules (***IRAK2***), and genes involved in the regulation of transcription (***NFKB2, NFKBIZ***).

Conclusion:

Genome wide gene expression provides a novel and useful means to investigate the molecular mechanisms of asthma phenotypes. Neutrophilic asthma has a distinct gene expression profile that is associated with upregulation of genes in the innate immune response.

Submitted by Katie Baines - katherine.baines@newcastle.edu.au

5

Analysing post-transcriptional gene control by next generation sequencing

Traude Beilharz, Jeff Squires, Meghna Sobti[1], Arthur Liu[1], David Humphreys[1], and Thomas Preiss[1,2]

1: Victor Chang Cardiac Research Institute;

2: University of New South Wales

Throughout their life in the cell mRNAs dynamically associate with many regulatory complexes, which ultimately determines their utilisation during protein synthesis by ribosomes. It is thus clear that a view of mRNA as merely a passive messenger or template is too simplistic and instead it is more appropriate to consider translation as a process of similar regulatory finesse as transcription. Consequently, many global techniques have been described to study gene regulation at the post-transcriptional level. We have been developing and employing several such techniques, initially using microarrays as our main experimental tool. More recently, we have gained access to an Applied Biosystems SOLiD 3 Next Generation Sequencing (NGS) system and are now transitioning our projects to this superior technology. For instance, we are currently developing methods to 'foot-print' translation factors and ribosomes along mRNA to arrive at entirely novel descriptions of the translation process. In these approaches, which are similar to the popular ChIP-Seq methods in DNA research, we aim to critically test the widely accepted closed-loop and scanning models of translation initiation. Another focus is the poly(A) tail of mRNAs, which serves as a tuneable promoter of translation and stability. We can now measure poly(A) tail lengths on a transcriptome-wide level using a combination of poly(U) chromatography and NGS. Given that that miRNAs commonly induce the deadenylation of their mRNA targets, we are now applying this approach to the discovery novel miRNA targets. Finally, we have been working on methods for the detection of 5-methylcytosine in RNA, by bisulfite treatment, and have started to combine this approach with NGS with a view of detecting novel sites of modification of RNA.

6

Pharmacogenomic approaches to the NCI60 database for breast cancer therapeutics

Tony Blick[1], Cletus A Pinto[1,2], Mark Waltham[1,2], Erik W Thompson[1,2]

1: St. Vincent's Institute, Melbourne

2: University of Melbourne Department of Surgery, St. Vincent's Hospital, Melbourne

Despite significant progress in the detection and treatment of breast cancer, and a corresponding increase in breast cancer survival over the last decade or two, better diagnostic and therapeutic options are required for certain breast cancer subtypes (e.g. basal breast cancers), for pre-existing metastases, and for recurrence.

We are interested in the development of new therapeutic possibilities against breast cancer and have developed an algorithm for examining the correlations across gene signatures with drug sensitivity data for the NCI-60 cell lines 1. The microarray and drug-sensitivity datasets are publicly available and this new algorithm is intended to compliment other pathway based analytical webtools currently available.

The algorithm is in the form of an Excel macro that reiteratively correlates each gene within a signature with a dataset of 4463 drug sensitivity profiles. The output is in the form of a cross-table ready for clustering, with R2 values for gene-drug pairs where the R2 value exceeds a user-defined cut-off.

We are in the process of applying the algorithm to datasets representing different breast cancer attributes.

1. Scherf, U., et al. (2000) Nat Genet 243, 236-244

7

Predicting Xeroderma Pigmentosum complementation group by gene expression profiling.

Nikola A. Bowden[1,2], Katie A. Ashton[1,2], Geoffrey Stibbard[1,2], Mathew B. Cox[1,2], Katherine Baines[1,2], and Rodney J. Scott[1,2,3]

1: Hunter Medical Research Institute

2: University of Newcastle

3: Hunter Area Pathology Service

Individuals with the autosomal recessive disease Xeroderma Pigmentosum (XP) have a 1000-fold increase in the risk of developing sunlight induced skin malignancies. XP is comprised of 7 complementation groups (XP-A to XP-G), which represent functional deficiencies in 7 genes involved in the DNA repair pathway, nucleotide excision repair (NER). The exact relationship between the different complementation groups remains to be precisely defined. Whole genome gene expression analysis (Illumina Ref8 V2) was utilized to identify similarities and differences after UVC-light exposure between fibroblast cell lines from the complementation groups XP-A to XP-G and unaffected controls. Gene transcripts with significantly altered expression and greater than 2 fold change were identified for controls and each XP complementation group. There was a subset of gene transcripts represented across all groups; however the control and XP complementation groups had feature sets of highly specific gene transcripts altered after UV irradiation. A graded change in gene expression patterns between the mildest NER deficiency (XP-E) and the severest NER deficiency (XP-A) was also indentified. The gene expression profile of an individual with an unknown XP complementation group was compared to the 7 XP gene expression profiles and the most likely complementation group was predicted based on the correlation of profiles and clinical characteristics of the individual. This analysis has revealed distinct gene expression profiles for each of the XP complementation groups and has provided the first step towards a viable alternative to complementation studies in the diagnosis of this disorder.

Submitted by Nikola Bowden - nikola.bowden@newcastle.edu.au

8

Challenges of assembling next-generation sequencing data generated from microbial enrichments

Lauren M. Bragg[1,2], Stuart E. Denman[1], Paul Evans[1], Glenn Stone[1], Jagadish Padmanabha[1], Andre D. Wright[1], Chris S. McSweeney[1], Dave Edwards[2], and Mark Morrison[1,3]

1: CSIRO Australia

2: University of Queensland

3: The Ohio State University, USA

Very few microbial genomes have been isolated in culture. During the era of Sanger-sequencing, attempting to sequence microbial genomes from a heterogeneous population was unfathomable.

Next-generation sequencing has provided a high-redundancy, low-cost means to sequence species which can not be isolated in culture. Sequencing of DNA with mixed microbial origin is commonly referred to as 'metagenomics', as the sequenced data represents the 'metagenome' of the microbial community. The new challenges presented by next-generation sequencing coupled with the complexity of metagenomic data significantly reduce the number of existing tools appropriate for analysis and assembly.

Here, we apply Illumina GAII sequencing (3kb-insert read-pairs) to a methanogenesis-favouring enrichment of the rumen contents from a Brahman steer (Bos indicus), and also from the foregut of the Western Grey Kangaroo (***Macropus fuliginosus***). Our primary aim is to obtain the genome sequence of uncharacterised archaeal methanogens. These methanogens are at approximately 50% relative abundance in both enrichments. The remaining 50% consists predominantly of members of the Clostridial and Bacteroides genera.

The insert-size distribution contained a mixture of long (3kb) and short (0.3Kb) inserts. As the majority were short-insert in both samples, all read-pairs were regarded as 0.3Kb insert library for the Velvet assembler. Expected coverage was estimated from a preliminary run of Velvet on both samples. Insufficient coverage was obtained for the methanogens in the Kangaroo sample.

Using the Brahman steer sample, Velvet assembled 931 contigs with an n50 of 53,562 and a maximum contig size of 425,511bp. After further in-house bioinformatics, we were able to scaffold the methanogen contigs into a circular molecule consisting of 1.46Mb, with only two significant predicted gaps, both less than 3kb.

Submitted by Lauren Bragg - lauren.bragg@csiro.au

9

The Children's Hospital at Westmead Tumour Bank: A dynamic biospecimen resource for translational cancer research

Albert Chetcuti, Nicole Mackie, Kerrie Jones, Guy Nelmes and Daniel Catchpoole

Tumour Bank, Oncology Research Unit, The Children's Hospital at Westmead, Westmead, NSW 2145 (www.chw.edu.au/tumourbank).

Access to ethically collected and well annotated tumour tissue is central to translational research. As a key resource for translational research, the Children's Hospital at Westmead Tumour Bank has developed into an important biospecimen repository for the international cancer research community. The aims of the Tumour Bank include: providing a link between samples collected for research and projects conducting translational research; establishment of protocols for handling research specific specimens; and implementation of research techniques that maximise the Tumour Bank's tissue resources.

The Tumour Bank focuses on collection, storage and distribution of specimens collected from paediatric cases treated at the Children's Hospital at Westmead. In addition, the Tumour Bank has implemented techniques and technologies that maximise the use of rare tumour specimens. These include; developing procedures for the preparation of DNA, RNA or protein; preparation of sectioned tissue; manufacturing of tissue microarrays; and using virtual microscope technology to scan sectioned tissue and perform quantitative image analysis.

To date the Tumour Bank has collected over 24,000 specimens from over 2,360 patients. The Tumour Bank has supported 46 research projects investigating diverse research areas. The Tumour Bank has also developed specific tissue microarrays and utilised virtual microscopy to digitally quantify proteins in histology sections.

The Tumour Bank plays an important role in supporting retrospective studies and is an import tool for future prospective studies. The Tumour Bank constantly endeavours to support and establish new research projects utilising both the specimens and the research tools available.

Submitted by Albert Chetcuti - albertc1@chw.edu.au

10

Identification of new proteins to determine invasion in favourable Wilms tumour

Albert Chetcuti[1], Safiye Aktas[2], Nicole Mackie[1], Celal Ulger[3,4,5], Gokce Toruner[3,4], Joseph Rinaggio[6], Anthony Galante[7], Bin Tian[8], Patricia Soteropoulos[3,7], Meera Hameed[9], Jane Carpenter[1], Marvin Schwalb[3,4], James Dermody[3,4], Mualla Alkan[3,4], and Daniel Catchpoole[1]

1. Tumour Bank, Oncology Research Unit, Kids Research Institute, The Children's Hospital at Westmead, Sydney, Australia.

2. Dokuz Eylul University, Institute Of Oncology, Department Of Basic Oncology, Izmir, Turkey.

3. Department Of Paediatrics, New Jersey Medical School, Newark, New Jersey, USA.

4. Institute Of Genomic Medicine, New Jersey Medical School, Newark, New Jersey, USA.

5. Department Of Biology, Faculty Of Science & Art, Adnan Menderes University, Aydin, Turkey.

6. Department Of Diagnostic Sciences, New Jersey Dental School, Newark, New Jersey, USA.

7. Center For Applied Genomics, Public Health Research Institute, Newark, New Jersey, USA.

8. Department Of Biochemistry, New Jersey Medical School, Newark, New Jersey, USA.

9. Surgical Pathology, New Jersey Medical School, Newark, New Jersey, USA.

Wilms tumour is the most common paediatric renal malignancy. Although survival rates are high, new prognostic parameters are required to understand invasion capacity and improve treatment methods. The aim of this study is to identify new proteins associated with Wilms tumour.

A total of 15 Wilms tumour cases were obtained from the Children's Hospital at Westmead Tumour Bank. Gene expression profiles of the samples were determined by Affymetrix HU-I33A GeneChip microarray analysis. A total of 14 genes showed differential expression and the genes EphB2, RIN1, MSX1 and Bcl2A1 were confirmed by quantitative PCR. These genes were immunohistochemically evaluated on archival tumour tissues. The expression was evaluated by scanning slides on an Aperio ScanScope scanner and digital analysis performed.

In normal kidney expression all four proteins were observed in tubulary epithelium. MSX1 has statistically significantly low expression in invasive cases compared to non-invasive cases (p=0.014). EphB2 expression was higher in invasive Wilms tumour cases than non-invasive cases (p=0.048). There were no statistically significant difference between non-invasive and invasive Wilms tumour for Bcl2A1 (p=0.140) or RIN1 (p=0.311).

Our results indicate that MSX1 may play a functional role in the invasive capacity of Wilms tumours. RIN1 is a downstream effector of RAS and Bcl2A1 functions as an anti-apoptotic protein. RIN1 and Bcl2a1 may have a functional role in tumourigenesis of Wilms tumour, but does not determine the invasive capacity in our Wilms tumour cases. EphB2 is an ephrin receptor and has higher expression in invasive cases.

Submitted by Albert Chetcuti - albertc1@chw.edu.au

11

Oxidative stress sensing in Saccharomyces cerevisiae by the cell cycle transcription factor Swi6p

Joyce Chiu[1], Carole M. Tactacan[1], Ruby C.Y. Lin[1,2], Merridee A. Wouters[3] and Ian W. Dawes[1,2]

1: School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney, NSW 2052, Australia.

2: The Ramaciotti Centre for Gene Function Analysis, University of New South Wales, Sydney, Australia.

3: Victor Chang Cardiac Research Institute, Sydney, NSW 2010, Australia.

Yeast cells begin to bud and enter S phase when growth conditions are favourable during G1 phase. When subjected to oxidative stress, cells arrest at G1 delaying entry into the cell cycle allowing repair of cellular damage. Hence, oxidative stress sensing is coordinated with the regulation of cell cycle. We identified a redox sensing cysteine residue in the cell-cycle regulator of Saccharomyces cerevisiae, Swi6p, at position 404. Mutation of Cys404 to alanine abolished the ability of the cells to arrest at G1 upon treatment by lipid hydroperoxide. Upon 5 min exposure to lipid hydroperoxide, Cys404 of Swi6p residue became oxidised indicating its sensitivity of cellular redox equilibrium. Microarray analysis study revealed that mutation of the Swi6p Cys404 to alanine led to no repression of G1-cyclins CLN1 and PCL1 and their induction resulted in the progression into S-phase despite the presence of oxidative stress. In conclusion, Swi6p serves as a molecular sensor of oxidative stress via its redox-sensitive residue Cys404 and inhibits entry into the cell cycle by suppression of G1-cyclin expression.

12

Gene expression profiling in Multiple Sclerosis

Mathew B. Cox[1] Nikola A. Bowden[1] Rodney J. Scott[1,2] and Jeannette Lechner-Scott[1,3]

1: Hunter Medical Research Institute and University of Newcastle, NSW, Australia

2: Hunter Area Pathology Service, New Lambton Heights, NSW, Australia

3: Department of Neurology, John Hunter Hospital, NSW, Australia

Multiple sclerosis (MS) is an inflammatory neurological disorder where inflammation within the central nervous system results in myelin degradation and axonal damage. Although the aetiology of the disease remains unknown, it is thought to be autoimmune in nature with both environmental and genetic factors involved in the pathogenesis of the disease.

To identify genes and pathways associated with MS, we performed whole genome gene expression analysis (Illumina HumanRef-8 V2) using PBMCs from 24 treatment negative MS patients, 13 patients receiving interferon-β therapy and 39 healthy controls.

Statistical analysis (GeneSpring GX V10) was used to identify differentially expressed genes ($p<0.05$, fold-change >2). Comparison of treatment naive patients compared to healthy controls showed an MS gene expression profile of 51 differentially expressed genes, the majority of which are involved in immune regulation/response or signalling pathways. Pathway analysis identified four genes differentially expressed in the plasminogen activating cascade, components of which have previously been associated with MS. The effect of interferon-β therapy in MS patients compared to healthy controls identified 106 genes differentially expressed, with only 15 genes in common with the MS gene signature. This indicates that the therapy appears to act by ameliorating the majority of differential expression which may be causing disease symptoms. The 15 genes in common with the untreated MS patients is composed mainly of genes with immunological or signalling functions.

The identification of an MS gene signature, and genes not affected by therapy, highlights potential areas of interest for further investigations into the complex pathogenesis of MS.

Submitted by Mathew Cox - mathew.b.cox@studentmail.newcastle.edu.au

13

The Hematopoietic Transcriptome Atlas

CA de Graaf, J Choi, T Baldwin, Sargeant TJ, Robinson A, Smyth GK, and Hilton DJ

Walter and Eliza Hall Institute of Medical Resesarch

Hematopoiesis is a tightly regulated process responsible for maintenance of blood cells during health, and that is able to boost production in response to illness or injury. It is controlled by transcription factor networks that are, in part, regulated by cytokines and another ligands binding to cell surface markers. To help us identify candidate regulators of differentiation, we have produced a comprehensive collection of gene expression data using Illumina Mouse WG microarrays. To date, we have collected data on 47 different blood cell types, which represent over 150 microarray samples sorted by flow cytometry, including a range of stem, progenitor and mature blood cell populations. We plan to expand this data with more diverse range of cells, including ES and induced pluripotent cells.

Using this data we have been able to follow changes in gene expression along hematopoietic differentiation, often observing gradual changes in gene expression as cells become lineage committed. We have also identified groups of genes with lineage specific expression patterns. In order to discover if these lineage specific genes have a functional role in differentiation we are undertaking an shRNA screen to a role for these genes in hematopoiesis.

Submitted by Carolyn de Graaf - degraaf@wehi.edu.au

14

High throughput oncogene mutation screening in melanoma cell lines

Ken Dutton-Regester[1,2], Lauren Aoude[1], Chris Scmidt[1], Adrian Herington[2], Nick Hayward[1]

1: Queensland Institute of Medical Research

2: Queensland University of Technology

INTRODUCTION:

The identification of gain-of-function mutations occurring in cancer has clinical applications for determining biological subgroups and assessing the efficacy of specific drug treatments. The ability to identify these mutations in a high throughput and accurate manner is desirable for use in a clinical setting.

METHOD:

The Sequenom OncoCarta v1.0 mutation panel was used to screen for oncogene mutations in stage III melanoma cell lines (n=40). The OncoCarta panel consists of 248 mutations in 19 oncogenes that is an optimised subset of those originally reported in high throughput screen of lung cancers. Mutant and wild type alleles are distinguished through the use of single base primer extension of amplified PCR products and detected via mass spectrometry using the Sequenom Mass Array platform.

RESULTS:

A total of 39 mutations were detected in 32 out of 40 melanoma cell lines. The majority of mutations detected occurred in BRAF (n=22 or 57.5%) and NRAS (n=9 or 22.5%) and were mutually exclusive consistent with current literature. A small number of BRAF mutated cell lines also had an additional mutation detected in CDK4, PIK3CA, HRAS or MET (n=2, 2, 2 and 1 respectively).

CONCLUSION:

The OncoCarta panel is an efficient method for high throughput and sensitive oncogene mutation detection. Refining the OncoCarta panel for biologically relevant assays specific to cancer subtypes would be a cost effective solution for mutation analysis in clinical applications.

Submitted by Ken Dutton-Regester - ken.dutton-regester@qimr.edu.au

15

Identification of miRNA expression profiles differentiating hepatitis c genotype 1 and 3 liver injury

WMH d'Avigdor[1], M Stapelberg[1], GW McCaughan[1,2], GW McCaughan[1,2], Maggie Lee[1], FJ Warner[1], NA Shackel[1,2]

1: Liver Laboratory, Centenary Institute, Sydney Australia

2: A.W. Morrow Gastroenterology and Liver Centre at Royal Prince Alfred Hospital, Sydney Australia

Introduction: Intrahepatic infection with the two most common HCV genotypes, 1 and 3 (G1 and G3), are characterized by poor interferon responsiveness and increased steatosis/insulin resistance respectively. Aim: to compare intrahepatic micro RNA (miRNA) expression in HCV G1 and G3 infection.

Methods: Applied Biosystems TLDA miRNA arrays compared HCV G1, G3 and control liver groups. miRNA expression was normalized to RNU48 and RNU6B expression. Intrahepatic injury was based on Scheuer and Metavir scores. Hierarchal clustering and Significance Analysis of Microarray data determined differences between groups.

Results: The miRNA expression of 9 HCV G1, 9 HCV G3 patients and 2 control non-diseased liver samples was compared. HCV G1 and G3 were split into equal groups of 3 mild and 3 severe intraheptic injury and 3 established cirrhotic samples (F4). Hierarchal clustering showed that mild and severe HCV injured liver is markedly different. Comparing HCV cirrhotic samples alone between genotypes, miR-146a, miR-223, miR-142-5p, miR-16 and miR-425-5p were significantly up regulated in G3. Irrespective of disease severity, miR-423 is significantly down regulated in G3 injured livers compared to G1 and control samples. Comparing cirrhosis, irrespective of genotype, to those individuals with minimal fibrosis (F0 or F1), miR-10a, miR-125a, miR-195a, miR-199a*, miR-200c and miR-214 were significantly up regulated and miR-192 and miR148a were down-regulated.

Discussion: Distinct miRNA expression is associated with HCV infection as well as HCV G1 and G3 liver injuries. The identified miRNA may regulate molecular pathways important in aspects of intrahepatic HCV pathogenesis, determining disease progression as well treatment outcomes.

Submitted by William d'Avigdor - williama@centenary.usyd.edu.au

16

Copy number variation in Childhood Acute Lymphoblastic Leukaemia

Tiffany-Jane Evans[1], Nikola A. Bowden[1], Bente A. Talseth-Palmer[1], Daniel Catchpoole[2], and Rodney J. Scott[1,3]

1: Hunter Medical Research Institute and University of Newcastle, NSW, Australia

2: The Tumour Bank, Oncology Research Unit at The Children\'s Hospital at Westmead

3: Division of Genetics, Hunter Area Pathology Service, John Hunter Hospital, Newcastle, NSW Australia

Acute Lymphoblastic Leukaemia (ALL) is the commonest childhood cancer in developed countries and accounts for approximately one third of all childhood malignancies. Although cure rates approach 80%, the aetiology of ALL is varied and largely undefined. A combination of genetic and environmental factors is believed to contribute to ALL development and there is significant interest in identifying alleles that are associated with this disease. The role of genomic copy number variation (CNV) is of growing interest in the identification of DNA duplications and deletions, and can disrupt normal gene expression thereby adversely affecting cell function. In this study, CNVs have been examined within a population of childhood ALL patients (n=99) who were treated for their disease and were, at the time of blood collection, considered to be in remission. The Illumina 370KCNV BeadChip arrays were used to estimate genome-wide changes in peripheral blood lymphocytes to gauge the extent and frequency of CNVs in this highly selected population. Using the visualisation tools GenomeStudio (Illumina) and Nexus (BioDiscovery) CNV profiles were produced for the cohort to identify common regions of variation and common pathways which are variably impacted by the observed CNVs. CNVs were found in or near a number of genes belonging to specific pathways known to be linked with ALL pathogenesis.

Submitted by Tiffany Evans - tiffany-jane.evans@newcastle.edu.au

17

Interferome: Integration of microarray datasets in the search for biological targets

Sam Forster, Shamith Samarajiwa, Katie Auchettl, Jodee Gould, and Paul Hertzog

Center for Innate Immunity and Infectious Diseases, Monash Institute of Medical Research, Monash University, Clayton, Vic., Australia

The Interferons (IFNs) are a group of potent antiviral cytokines intimately associated with immune response. Since discovery in 1957 numerous cell and tissue specific functions have been identified and attributed to interferon signaling. These include protection against viral and bacterial infection and regulation of cell proliferation, differentiation, survival, migration and development and immune cell functions. Each type and subtype of IFN demonstrates unique tissue, cell and condition specific expression both basally and in response to pathogen infection. IFNs are used as therapeutics in many diseases such as chronic viral infections, cancer and multiple sclerosis. Naturally occurring, aberrant IFN signaling is also associated with many diseases particularly inflammatory and autoimmune diseases.

The INTERFEROME, was conceived as an integration of interferon stimulated, high-throughput microarray datasets resulting in a complete list of genes regulated by IFNs. These genes, known as IFN regulated genes (IRGs) were collected through analysis of over 40 in-house and published IFN microarray and proteomic datasets from 5 species including humans, mice and chimpanzees. This study, published in early 2009 resulted in identification of approximately 2000 genes with transcription regulated by IFNs in a subtype, dose, cell type and time dependent manner. In addition to microarray data this database, freely available at http://www.interferome.org/, also facilitates identification of regulatory elements, tissue expression patterns, comparative genomic and phylogenetic analysis across 37 species.

The INTERFEROME provides a central resource for infection, immunity and cancer research assisting in the identification of IFN signatures within gene lists generated through high-throughput expression technologies. With the ongoing integration of further datasets and development of pathway and functional network information we have been able to predict mediators of cancer metastases and viral infection evasion mechanisms

Submitted by Sam Forster - sam.forster@med.monash.edu.au

18

Detecting changes in chromatin state over time.

Hugh J French, Kristine Hardy, M Frances Shannon, Rohan BH Williams, and Kaiman Peng

The John Curtin School of Medical Research, The Australian National University, Canberra, Australia

T cells are a central component of the immune response. The activation of T cells is an important step in this response. Activation of T cells causes wide-spread changes in their global gene expression. The basal chromatin state of genes influences the kinetics of gene induction. Chromatin Immunoprecipitation combined with DNA tiling microarrays provide a means to detect histone modifications across the entire genome. Utilising these genomewide approaches, changes of chromatin state of genes upon T cell activation may be observed. The influence of these changes on inducibility of transcription maybe then inferred. Here we present methodologies to detect genes and groups of genes that undergo changes in histone modification status across T cell activation.

Submitted by Hugh French - hugh.french@anu.edu.au

19

Defining MYB transcriptional network

Evgeny A. Glazov, Liang Zhao, Paul Leo, Ping Zhang, Diwakar R. Pattabiraman, Matthew A. Brown, and Thomas J. Gonda

Diamantina Institute for Cancer, Immunology and Metabolic Medicine, University of Queensland, Brisbane, Queensland, Australia

MYB is a transcription factor required for maintenance of progenitor and stem cell populations in the bone marrow, colonic crypt and neurogenic niches. MYB has also been identified as an oncogene involved in several common cancers including haematopoietic malignancies, colorectal cancer and breast cancer.

On the molecular level MYB operates predominantly as a transcriptional activator through binding to MYB binding site consensus DNA sequence (t/cAACt/gG). Although over 80 target genes of the MYB transcription factor have been identified, this is not sufficient to explain some important aspects of MYB action, such as ability to block differentiation. Moreover, many of these targets were identified in RNA expression profiling experiments that can't always differentiate between genes directly affected by MYB and indirect secondary changes in gene expression.

To identify a comprehensive set of genes directly targeted by MYB we used chromatin immunoprecipitation combined with ultrahigh throughput sequencing (also known as ChIP-Seq) in conditional MYB transformed mouse myeloid progenitor cells (ERMYB). We identified over 4000 thousands high confidence MYB binding sites in mouse genome. This approach allowed us to validate most of known MYB target genes and identify large number of new ones including several miRNA genes. It also enabled us to refine MYB binding consensus sequence and to uncover correlation between strength of MYB binding and preferred binding site sequence composition. In addition, we found a significant overrepresentation of several other transcription factor binding sites including SP1, and ETS transcription factor family members suggesting regulatory interplay between MYB and these transcription factors.

Submitted by Evgeny Glazov - e.glazov@uq.edu.au

20

Visualising time-series microarray data on the interactome

Apurv Goel, Simone Li, Chi Nam Ignatius Pang, Marc R. Wilkins

Systems Biology Initiative, School of Biotechnology and Biomolecular Sciences, University of New South Wales

A data analysis platform has been built upon a 3D visualisation engine (GEOMI), allowing the covisualisation of time-series gene expression data, protein-protein interaction networks and many other aspects of proteins (e.g. localisation, abundance, biological process). First, data from time-series microarrays are plotted and smoothed to reduce errors and fill in any missing time points. This smoothed plot is then applied to an existing protein interaction network, whereby nodes (representing proteins) are coloured depending on their changes in expression. The result is a looping animation, where node colour changes in real-time to match expression values in the experimental time. The colour scheme that is used is the same as that used in microarrays; heatmaps. Red indicates overexpression, green indicates underexpression and black indicates no change in expression. This visualisation method is not only limited to time-series data, any series-based data can be analysed (e.g. changing concentrations of stressor). Conceptually, however, it makes the most sense on time-series data since the time axis is being mapped to the 4th dimension.

This method of visualisation is providing a deeper insight into the dynamics of the interactime. The resulting visualisations can be used to quickly determine expression patterns for proteins and how they relate to interacting proteins over time or in different environmental conditions. It also allows the identification of co-regulated or differentially regulated proteins in complexes. GEOMI can be downloaded at www.systemsbiology.org.au.

Submitted by Apurv Goel - apurvgoel@gmail.com

21

Where do carcinoma-associated fibroblasts come from?

Robin Anderson[1], Wen Qiu[1], Ian Campbell[1,3], Adam Kowalczyk[2,3], Christina Restall[1], Richard Tothill[1], Alex Boussioutas[1], David Bowtell[1,3], Terry Speed[4], Min Hu[5], and Kornelia Polyak[5]

1: Peter MacCallum Cancer Centre

2: National ICT Agency

3: University of Melbourne

4: Walter and Elisa Hall Institute

5: Dana Farber School of Medicine, Harvard University

Gene expression analysis of a whole tumor sums the mRNA contributions from different cell lineages, driven by cell-autonomous genomic changes as well as non-cell autonomous cell-matrix and cell-cell interactions. Here we describe a microarray analysis of defined co-cultures to identify intercellular signals that contribute to cross-talk between ovarian epithelial cells and fibroblasts. A reciprocal interaction was found, where fibroblast-derived IL8 and EGF provoke the epithelial cells to express TNFx, thus driving an inflammatory gene expression signature in fibroblasts that is recapitulated in a robust and tightly co-expressed signature of gene in human cancer expression profiles. Additional genes in this signature from human expression profiles reflect infiltrates of inflammatory cells that respond to the fibroblast-derived chemokines, particularly CCL2, CCL5 and CCL8. A comparative semi-supervised analysis based on clinical outcome confirmed that the expression of this gene signature is exhibited in a group of patients with inherently poor prognosis, due to elevated metastatic potential, and microdissection of tissue sections showed that indeed these genes are in fact expressed in tumor-juxtaposed fibroblasts. In animal models of prostate, lung, ovarian and breast cancer, normal epithelial cells can become malignant when surrounded by fibroblasts that are derived from tumors (Carcinoma-Associated Fibroblasts). We therefore seek the molecular basis of this phenotypic carcinoma promotion of fibroblasts within cancer tissue. In spite of the robust reaction of fibroblasts and epithelial cells to co-culture, in our hands, fibroblasts from normal reduction mammoplasty fail to turn into cancer promoting fibroblasts via co-cultivation in xenografts with cancer cells (see preliminary results). Furthermore, fibroblast genome was found to be intact in 21 of 22 tumor-juxtaposed fibroblasts obtained through microdissection of ovarian and breast cancer sections. We demonstrate that mesenchymal progenitors (MP) infiltrate primary tumors. Effective recruitment of the MP to primary tumors, to assist in tumor cell dissemination and/or the establishment of metastatic sites depends on specific genes, such as CCL5, but the recruiting signal from the cancer remains elusive. This work offers a potential source for this recruitment, i.e. the residential fibroblast transient response to tissue injury and interaction with the cancer cell. In addition, this work offers a compendium of experimental growth conditions and their consequential gene expression changes, and assesses which of those are most robust and consistent in a large fraction of cancer patients of a variety of carcinomas. The resultant interaction-expression associations may focus future attempts to modulate the cancer microenvironment.

Submitted by Izhak Haviv - izhak.haviv@bakeridi.edu.au

22

Altered transcriptional responses in elite endurance athletes exposed to moderate altitude

Jennifer Henderson[1], Vicky Cho[1], Laura Garvican[2,3], Bon Gray[4], Chris Gore[2], Rohan Williams[1], Simon Easteal[1]

1: John Curtin School of Medical Research

2: Australian Institute of Sport

3: Flinders University

4: Bond University

The cellular response to reduced oxygen tension is ubiquitous among organisms and is largely mediated by hypoxia-responsive activation of gene expression. The extent of inter-individual variation and the specificity of transcription induced in response to hypoxic stress in humans and how this relates to healthy adaptation remain unclear. Exposure to local or systemic hypoxia is considered a key signal for many adaptations such as those associated with endurance exercise. Athletes commonly live and train for a period of time at moderate terrestrial altitude (~1,500- 3,000 m) to improve sea level performance. The response is highly variable and often the expected physiological and/or performance benefits do not materialize. We tested the hypothesis that the response to altitude exposure may be mediated by changes in global gene expression by assaying gene expression (real-time PCR and microarrays) in a cohort of elite endurance cyclists undergoing a controlled 3-week training camp at either moderate altitude or near sea level. We present evidence of an attenuated transcriptional response in elite cyclists training at moderate altitude in genes normally activated by hypoxia. The time-course and nature of the transcriptional response observed was highly variable and may be dependent on baseline levels in selected transcripts. The results suggest long-term endurance adaptation may lead to an acquired hypoxic tolerance that occurs in the absence of transcriptional activation of traditional hypoxia-responsive genes. Further research is required to elucidate how altitude exposure alone or compounded by elite endurance training acts to repress gene expression in normally hypoxia-responsive genes in endurance-trained athletes.

This research was partially funded by the Australian Institute of Sport, the NSW Institute of Sport & Bond University.

Submitted by Jennifer Henderson - jennifer.henderson@anu.edu.au

23

Differential variability of gene expression in human diseases and ageing

Joshua W.K. Ho[1,3] and Michael A. Charleston[1,2]

1: School of Information Technologies, The University of Sydney

2: Sydney Bioinformatics and Centre for Mathematical Biology, The University of Sydney

3: NICTA, Australian Technology Park, NSW

Many microarray gene expression studies are designed to investigate the steady-state gene expression changes between different sub-populations, such as identifying genes that are differentially expressed (DE) or differentially coexpressed (DC) in individuals which have different phenotypes (e.g., diseased vs. non-diseased, or ageing). By re-analyzing a number of public microarray datasets related to human diseases and ageing, we found genes that have significant alteration in expression variability (variance) between different phenotypic groups. Both increase and decrease in gene expression variability are possible, but it is more common to find genes that have elevated expression variability in diseased or aged individuals. Moreover, we show that genes with DV tend to also be differentially coexpressed, which implies that loss of expression variability may be associated with a loss of gene regulation. It is possible that the observed change in expression variability is caused by varying rates or mechanisms of gene regulation among individuals in different phenotypic groups. Therefore identification and characterization of such genes are of importance in understanding why certain genes are responsive to external perturbation (such as medication) in some individuals while the same perturbation yields little response in other individuals. In this talk, we will present novel applications of statistical methods to discover genes with DV from microarray gene expression profiles, as well as presenting analysis results from various human diseases and ageing microarray datasets. We hope that our work can provide new insight into how population based transcriptomic data can be analyzed in the era of systems biology and personalized medicine.

Submitted by Joshua Ho - joshua@it.usyd.edu.au

24

Gene expression analysis of colorectal cancer subtypes based on the Jass system

Y. Hu and G. K. Smyth

Walter and Eliza Hall Institute of Medical Research

Colorectal cancer has not so far yielded a clear clustering of tumours into subtypes based on gene expression profiles in the way that breast and ovarian cancer have done. Instead, the Jass System is one of the most accepted classifications. This system classifies colorectal cancer into five subtypes, situated around a continuum, based on a combination of clinical, morphological and molecular features. In this study, we dissect the gene expression of profiles of the five Jass subtypes, using data on 45 patients from the Colon Cancer Family Registry.

We demonstrate the important of correcting for batch effects and allowing for variable quality between the RNA samples in the analysis. Pair-wise comparisons were made among 5 tumour classifications while down-weighting poor quality samples and including random effects for sample hybridization day. Differential expression analysis shows that the tumours divide into two major classes based on the molecular profiles, and these correspond broadly to distinct mutated pathways.

Next we developed a novel gene set enrichment analysis method suitable for a study of this type with many subgroups, batch effects and quality weights. We apply this method to interrogate the tumour subtypes using a battery of co-regulated gene sets provided by the Broad Institute's Molecular Signatures Database. This yields many interesting pathways and sets of co-regulated genes which distinguish the groups.

This project is a collaboration with the National Cancer Institute's Colon Cancer Family Registry and the Australasian Colorectal Cancer Family Study.

Submitted by Yifang Hu - hu@wehi.edu.au

25

Next-generation sequencing of the molecular events in breast cancer

K.Krishnan[1], N.Cloonan[1], C.Wong[2], A.Moller[2], and S.Grimmond[1]

1: Expression Genomics Laboratory, Institute for Molecular Bioscience, University of Queensland, St. Lucia, Queensland.

2: Cancer Genomics & Biochemistry Laboratory, Peter MacCallum Cancer Centre, East Melbourne, Victoria.

Cancer is a debilitating disease driven by the accumulation of various genomic and epigenomic alterations, much of which are reflected in the transcriptome. Though CGH and microarray studies enable identification of some of these mutations, we still need to characterize the cancer genome and transcriptome at single nucleotide resolution to elucidate its underlying molecular events.

To achieve this, we propose to use the SOLiD next-generation sequencing technology, coupled with a spontaneous breast cancer transgenic mouse model, Polyoma Middle-T Oncoprotein (PyMT), which closely recapitulates classical markers of the human disease.

In this study, we have sequenced the genomic DNA of a PyMT tumour and identified structural and copy number variations (SVs & CNVs). The RNA from the same tumour has been sequenced and the data has been annotated in our pre-existing pipeline which quantitates gene activity, splice variant expression and sequence variation analysis. Libraries of DNA enriched for methylation are being generated.

Genes differentially expressed between normal and cancer samples correlating with genomic aberrations and/or differential methylation patterns will be picked as candidate tumour suppressors/ activators. The mutations will first be confirmed using conventional techniques. Once validated, their relevance to human disease will be assessed and ranked using publically available CGH and microarray data. The highly ranked genes will be screened in an independent panel of human tumour samples and followed up with functional studies.

We hope this approach will allow unbiased and rapid discovery of novel oncogenes and tumour suppressors, leading to novel biomarker discovery and potential development of new therapeutics.

Submitted by Keerthana Krishnan - k.krishnan@imb.uq.edu.au

26

Bioinformatic evidence for transcription factor interactions: AML1 with ERG and ETS2

Charity Law, Catherine Carmichael, Wei Shi, Gordon K Smyth

The Walter and Eliza Hall Institute of Medical Research

Transcription factors are believed to interact with one another in complex ways, although only a small proportion of these interactions are well described. Here we study bioinformatically the interactions between three transcription factors, AML1, ETS2 and ERG, known to play an important role in a range of cancers.

The transcription factors AML1, also known as RUNX1, and CBFB are known to interact with each other, forming a heterodimeric complex that is more stable and more able to bind DNA. AML1 is also known to interact with other transcription factors such as ETS1 and NERF-1/2, which belong to the ETS family of transcription factors. However no interactions between AML1 and ETS2 or ERG, which are also a part of the ETS family of transcription factors have so far been established.

Here we explore the evidence for interaction between the AML1/CBFB-complex with ERG and ETS2 using microarray data and binding site data. Two-colour microarray data is used with the overexpression of transcription factors individually and in combination. Various linear modeling techniques are used as evidence for the presence of interaction between these transcription factors.

27

Caution! Compositions! Can constraints on omics data lead analyses astray?

David Lovell[1], Warren Müller[1], Jen Taylor[2], Alec Zwart[1]

1: CSIRO Mathematical and Information Sciences

2: CSIRO Plant Industries

Some DNA or RNA sequencing methods produce data that can be considered as counts of the number of times each sequence was observed in the biological sample. Because the sum of these counts within each sample is constrained by the sequencing process and the physical amount of the sample, they constitute compositional data (Aitchison, 1986). There are many other examples of compositional data in the omics, including relative abundances of species (in metagenomics) or GO terms (in functional genomics). Few researchers have broached the issue of analysis of compositional data in omics count surveys, but in the geosciences there has been debate for nearly half a century about how sum-constrained data should be analysed. This presentation aims to raise awareness of whether, and in what circumstances, naïve analysis of sum-constrained data could lead to incorrect inference, and explore the extent to which this might be a problem in omics applications.

Submitted by David Lovell - David.Lovell@csiro.au

28

Biological replicate analysis of one-colour Affymetrix microarrays

Victoria Lyons, Gal Winter, Patrick J O'Doherty,Ming J Wu, and Vincent J Higgins

School of Biomedical and Health Sciences, Campbelltown Campus, University of Western Sydney, Locked Bag 1797, Penrith South DC, NSW 1797, Australia

Yeast genome-wide differential gene expression can quickly and easily be assessed using microarrays such as the one-colour Affymetrix chips. The high-throughput nature of microarrays and the immense amount of data generated has led to the scientific community generally expecting verification by qPCR and further experiments before publication of such data. If this is the case, triplicated microarray data may not be necessary for these projects because similar results can be achieved using microarrays in duplicate, resulting in lower costs involved in microarray experiments. To verify this, microarray data that had been carried out in biological triplicate was analysed in a duplicate manner and the generated lists of genes were compared. Comparison of ANOVA analyses of the triplicate dataset and each of the three possible duplicate datasets showed R2 values of at least 0.95, indicating high correlation between these sets. Out of those genes that were significantly up- or down-regulated (having at least a 2-fold change), at least 68% were common between the triplicate gene list and each of the duplicate sets, with the genes that weren't common being those with minor differential gene expression. This suggests that lower eukaryote, yeast, microarray experiments can be carried out in biological duplicates, saving a-third of the costs involved without significantly changing the results obtained.

Submitted by Victoria Lyons - v.lyons@uws.edu.au

29

Independent amplification of 20q11.21: support for an oncogene in a 250kb region in acute myeloid malignancy with deletion of 20q12

Ruth N. MacKinnon[1] and Lynda J. Campbell[1,2]

1: Victorian Cancer Cytogenetics Service, St Vincent's Hospital, Melbourne

2: Department of Medicine (St Vincent's), University of Melbourne

We have identified a series of five cases of myelodysplastic syndromes (MDS) and acute myeloid leukaemia (AML) with loss of the 20q12 tumour suppressor gene region and localised amplification at chromosome band 20q11.21.

All five cases had gain of the same overlapping 250kb region containing four complete genes, one of which we suggest is an oncogene active in MDS/AML. Positive selection of a common 1Mb region occurred at least twice in one of these patients, evidence in support of an oncogene in this region.

In this case of AML, clonal evolution of the karyotype at relapse showed a narrowing of the extent of amplification of 20q. The 38 year-old male presented with erythroleukaemia and a dicentric (17;20) unbalanced translocation as the basic chromosome abnormality, together with more complex derived cell lines. At diagnosis array CGH showed there was an average gain of one copy of a 2Mb section of 20q11.21. At relapse different derivatives of the dic(17;20) predominated. There was an average gain of two copies of a 1Mb region from within the original 2Mb region, localizing around the 250kb shortest region of overlap of amplification. FISH confirmed these results and showed that there was a high level of amplification of this region in some cells.

These five patients had either a diagnosis of erythroleukaemia (AML-M6) or MDS with erythroid hyperplasia and morphology approaching erythroleukaemia. Our observations support a role for the acquisition of extra copies of a critical 20q11.21 region in the transformation of MDS with del(20q) to erythroleukaemia.

Submitted by Ruth MacKinnon - ruth.mackinnon@svhm.org.au

30

Disease-specific, neurosphere-derived cells as models for brain disorders

Nicholas Matigian[1*], Greger Abrahamsen[1,*], Ratneswary Sutharsan[1], Anthony L Cook[1], Amanda Nouwens[2,3], Bernadette Bellette[1], Alejandra M Vitale[1], Jiyuan An[1], Matthew Anderson[1], Anthony G. Beckhouse[1], Maikel Bennebroek[1], Rowena Cecil[1], Alistair M Chalk[1], Julie Cochrane[1], Yongjun Fan[1], François Féron[1],+, Richard McCurdy[1],++, John J. McGrath[1,4,5], Wayne Murrell[1], Chris Perry[1,6], Jyothy Raju[1], Sugandha Ravishankar[1], Peter A. Silburn[1,7], Greg T. Sutherland[1], Stephen Mahler[2], George D. Mellick[1], Stephen A Wood[1], Carolyn M. Sue[1,8], Christine A. Wells[1], Alan Mackay-Sim[1],

Affiliations

1: National Centre for Adult Stem Cell Research, Eskitis Institute for Cell and Molecular Therapies, Griffith University, Queensland, Australia

2: Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, Queensland, Australia

3: School of Chemistry and Molecular Biosciences, The University of Queensland, Queensland, Australia

4: Queensland Centre for Mental Health Research, The Park Centre for Mental Health, Queensland, Australia

5: Queensland Brain Institute, The University of Queensland, Queensland, Australia

6: Department of Otolaryngology Head and Neck Surgery, Princess Alexandra Hospital, Queensland, Australia

7: School of Medicine, The University of Queensland, Brisbane, Queensland 4072, Australia

8: Kolling Institute of Medical Research, The University of Sydney, Royal North Shore Hospital, St Leonards, New South Wales, Australia

Patient-derived stem cells would be invaluable for understanding the molecular bases of nervous system diseases. The human olfactory mucosa is continually regenerating throughout life from multipotent neural stem cells which are accessible and expandable in vitro as neurospheres containing stem/progenitor cells. Here we demonstrate that patient-derived olfactory neurosphere-derived cells reveal disease-specific alterations in cell biology in a neurodevelopmental, psychiatric disorder (schizophrenia) and in a neurodegenerative disease (Parkinson's disease). Comparison of 42 olfactory cell lines from patients and controls demonstrated significant disease-specific alterations in gene expression, protein expression and cell function including dysregulated neurodevelopmental pathways in SZ and dysregulated mitochondrial function, oxidative stress and xenobiotic metabolism in PD. The cells revealed new candidate genes and cell pathways for future investigation. Fibroblasts from SZ patients did not show these differences. Olfactory stem/progenitor cultures provide an alternative to iPS and ES cells as disease models. They do not require genetic re-programming and they can be obtained from adults with complex genetic diseases. They will be useful for understanding aetiology, for diagnostics and for drug discovery.

31

The impact of technical replicates on the interpretation of single-channel Affymetrix microarrays

Patrick J. O'Doherty, Ryan A. Hyland, Victoria Lyons, Ming J. Wu, and Vincent J. Higgins

School of Biomedical and Health Sciences, Campbelltown Campus, University of Western Sydney, Locked Bag 1797, Penrith South DC, NSW 1797, Australia.

Differential gene expression analysis using microarray technology has become a standard investigative tool in molecular biology. To achieve optimal results statistically from microarrays, it has been advocated that replicates be included in the experimental design. However, it is widely accepted that technical replicates are not necessary when biologically independent datasets sampled in triplicate are analysed using single-channel microarrays. To verify whether this is actually the case, single-channel microarray analysis consisting of a technical duplicate of a biological triplicate was carried out. Comparison of the ANOVA analyses between the two technical sets of biological triplicates showed an R2 value of 0.9427, indicating high correlation between these datasets. Of the genes significantly up-regulated (minimum of 2-fold change) within these data sets, 82.4% showed commonality, likewise, of the significantly down-regulated genes 80.7% were common. The remaining genes that were not common were of minimal fold change. This suggests that technical replicates do not significantly contribute to the results of single-channel Affymetrix gene chip analysis for experiments performed in biological triplicate.

Submitted by Patrick O'Doherty - p.odoherty@uws.edu.au

32

Application of gene set testing to probe regulatory T cell activity in gene expression data from early multiple sclerosis patients.

V.M. Perreau[1], O. Skibina[2], Y. Hu[3], M. Binder[4], J. Field[4], H. Butzkueven[4], G.K. Smyth[3], and T. J. Kilpatrick[1,4]

1: Centre for Neuroscience, The University of Melbourne, Parkville, VIC 3010, Australia.

2: Royal Melbourne Hospital, Parkville, VIC 3050, Australia

3: Walter and Eliza Hall Institute, Parkville, VIC 3052, Australia

4: Florey Neurosciences Institute, Parkville, VIC 3010, Australia

Although the cause of multiple sclerosis (MS) is unclear it is generally accepted to be an autoimmune disease with chronic inflammation in the central nervous system caused by autoreactive effector T cells crossing the blood brain barrier. T regulatory cells (Tregs), which comprise only 2-3% of total human CD4+ T cells, protect against autoimmunity by actively suppressing autoreactive T cells and there is increasing evidence that activation and maintenance of the Treg response could be dysfunctional in patients with MS.

We hypothesise that inflammatory changes involving Tregs occur in the earliest stages of multiple sclerosis and are reflected in the gene expression profile (GEP) of peripheral T cells, which are likely to be characteristic of levels of disease activity.

GEPs from CD3+ T cells were obtained from 11 patients and matched controls (collected at the time of presentation of first symptom (0-month) and 3 months later). Patients were segregated into 2 groups based upon presence (active group N=5) or absence (inactive group N=6) of disease activity at 3 months following presentation. The ROAST gene set test was applied using the limma software package to focus analysis on expression signature changes related to Tregs, exploiting the coordinate behavior of published lists of genes characterizing the expression signature of Treg activation and also that of expression of the ***FOXP3*** gene, a master regulator of Treg activity. We found that expression signatures of Treg activity, FOXP3 activation and cytotoxic T cell activity differ between these 2 clinically separable patient populations; active and inactive.

Submitted by Victoria Perreau - vperreau@unimelb.edu.au

33

CGH Microarray as a diagnostic tool in routine testing for genomic disorders: problems and prospects.

Greg Peters, Artur Darmanian, and Dorothy Hung

Cytogenetics Dept, Childrens Hospital at Westmead, Sydney, Australia

Many diagnostic laboratories are adopting microarray technologies as an alternative to the traditional method of genome-wide scanning: by karyotype. Our own laboratory began routine CGH array testing in 2006. In a select few cases, we have also used genotyping (SNP) arrays.

Whereas karyotype analysis offered a mean resolution of ~5Mb, microarrays can achieve 0.1 Mb or better:- although not without some drawbacks.

Here we consider the evolving applications of microarray technology in the diagnostic setting. Introduction of CGH array has led to a fourfold increase (at least) in detection of pathological change in gene copy number, in cases of paediatric developmental delay, +/- dysmorphism. However, this improvement in diagnosis has come with an increase in the detection of likely "false positives", in the form of innocuous polymorphic variants for sequence copy number.

In the long run though, accumulation of data on these non-pathogenic (neutral?) variants can also be helpful. For example: Such data may help us appreciate the extent to which there is selective constraint on deviation from disomy, in gene-dense vs gene-poor regions across the genome.

Submitted by Greg Peters - gregoryp@chw.edu.au

34

SNP-MaP of the severity and duration of acute illness.

B.F. Piraino[1], A.R. Lloyd[1], and U. Vollmer-Conna[2] on behalf of The Dubbo Infection Outcomes Study investigators.

1: Centre for Infection and Inflammation Research, School of Medical Sciences, University of New South Wales, Sydney, Australia

2: School of Psychiatry, University of New South Wales, Sydney, Australia.

Background:

Acute infective illnesses are associated with stereotyped symptoms, including fevers, musculo-skeletal pain, anorexia, and headaches. This response is believed to be an adaptive mechanism to promote recovery. It has both immunological and neurobehavioural components, and a wide spectrum of severity and duration. Single nucleotide polymorphisms (SNPs) in host response genes are known to contribute to the outcomes of infectious diseases.

We hypothesised that SNPs in immunological and neurobehavioural genes are associated with variations in the severity and duration of acute infection.

Subjects and methods:

DNA and clinical data were available from 400 subjects enrolled in the Dubbo Infection Outcomes Study – a prospective cohort examining outcomes from Epstein-Barr virus, Ross River virus or Q fever infections. A composite illness severity measure was derived from a principle components analysis of self report symptom data. Illness duration, a period from symptom onset to recovery, was derived for all subjects. DNA from subjects comprising the extremes of the severity and duration phenotypes (top and bottom thirds) were used in a genome wide screen employing the SNP Microarrays and pooling (SNP-MaP) approach using Affymetrix GeneChip® 100K arrays.

Results:

The SNP-MaP analysis revealed that 17 SNPs associated with illness severity (Bonferroni corrected $p < 4.32 \times 10^{-7}$) and 48 SNPs were associated with illness duration (Bonferroni corrected $p < 4.34 \times 10^{-7}$). Nine SNPs from each phenotype were confirmed by genotyping of individual DNA samples.

Conclusions:

Multiple genetic polymorphisms contribute to the severity and duration of illness in acute infection.

35

Bioinformatic analysis of merged gene expression microarray datasets in rat models of colorectal cancer.

Janet Shaw[1,2], Caroline A. Kerr[1,2], Julie Clarke[1,4], Trevor Lockett[1,2], Richard Head[1]

1: CSIRO Preventative Health Flagship CSIRO

2: CSIRO Food and Nutritional Sciences, North Ryde, NSW

3: CSIRO Food and Nutritional Sciences, Adelaide, SA, Australia.

Azoxymethane (AOM) induced carcinogenesis in rats is a central model used in the study of colorectal cancer (CRC). This analysis merged in silico gene expression data from two AOM studies that used different generations of Affymetrix GeneChip® Rat arrays. The first sampled rat colonic epithelium immediately after AOM injection (Acute model, Genome 230 2.0 Arrays), the second sampled colonic epithelium and AOM-induced tumours (Cancer model, 27 weeks after AOM exposure, Gene 1.0 ST arrays). Saline injected controls were included in both studies. The aim of this analysis was to compare the acute and long-term effect of AOM in the carcinogenesis process by merging these two datasets from different generation arrays.

The merger and gene expression analysis was performed using Partek® and pathway and network analysis using Ingenuity® Systems. Genes unique to either array were first filtered by Gene Symbol annotated probe-sets and then by CEL file. The resulting data was then normalised with RMA using the quantile distribution of the Acute arrays. As the Gene 1.0 ST array covers only well-annotated content, the resulting merged list consisted of 12,435 genes. Of this list, 11,437 (92%) were differentially expressed between the Acute and Cancer studies regardless of treatment. This large 'experimental effect' due to chip and/or 'rat' (e.g. age) differences was then modelled and the residuals analysed for differential expression between treatment groups.

Lists of differentially expressed genes comparing AOM treated rats to normal in proximal and distal colon epithelium were generated using an FDR of 0.05. These were then analysed for affected Functions and Pathways. Comparison analyses between the acute and cancer models found quite different Functions and Pathways involved in the acute compared to the long-term response of the colon epithelium to the AOM insult.

Submitted by Jan Shaw: Jan.Shaw@csiro.au

36

Cross platform integration of microarray data

Chintanu Kumar Sarmah[1], Sandhya Samarasinghe[1], Don Kulasiri1, Daniel Catchpoole[2]

1: Lincoln University (New Zealand)

2: The Children's Hospital at Westmead (Australia)

Gene expression profiling is rapidly evolving into a powerful technique for investigating tumor malignancies. The researchers, however, become overwhelmed with the Microarray-based platforms and methods that confer them the freedom to conduct large-scale gene expression profiling measurements. Therefore, investigations into cross-platform integration methods have recently started gaining momentum to get the maximum benefit from generally small number of experiments conducted on different platforms.

In this paper, we explain an adaptation of a recently published method of cross platform data normalization, XPN to childhood leukaemia gene expression analysis. The method, applied previously to Affymetrix and Agilent datasets, is applied with relevant modifications to cDNA and Affymetrix datasets. Artificial neural networks have been used in the process to investigate as well as retrieve the predominant attributes of the data, and are subsequently integrated into the method.

Keywords: gene expression profiling, microarray, cross-platform, cDNA, Affymetrix, Agilent. XPN, Artificial neural networks.

Submitted by Chintanu Sarmah - sarmahc3@lincoln.ac.nz

37

Estimating the proportion of microarray probes expressed in an RNA sample

Wei Shi and Gordon K Smyth

The Walter and Eliza Hall Institute of Medical Research, Melbourne

A fundamental question in the microarray analysis, which has yet been addressed, is the estimation of the number of expressed probes in an RNA sample. Detection calls have been used by many microarray studies to select a subset of probes on the array for further analysis. However, detection calls do not provide an estimate of the number of expressed probes.

Negative control probes which become available in the latest microarray platforms, such as Illumina whole genome expression BeadChips, provide a unique opportunity to estimate the number of expressed probes by taking advantage of their ability to measure the background noise. A model, for the first time, was proposed in this study to estimate the number of expressed probes in an RNA sample by utilizing these negative controls with no need to set a threshold.

Using the proposed model, the newer versions of Illumina BeadChips are found to have larger percentages of expressed probes than older versions, reflecting improved probe design in newer versions. Probes designed to interrogate the well-characterized RefSeq NM transcripts have a larger percentage of expressed probes than other probes in the array. RNA samples mixed from two pure samples are found to have more expressed probes than the pure samples, due to the fact that mixed samples have more distinct transcripts than any pure sample.

The proposed model has a variety of applications such as estimating the transcriptome size, filtering out non-expressed probes, improving array normalization and so on.

Submitted by Wei Shi - shi@wehi.edu.au

38

Molecular profiling of archival breast cancer samples using DASL technology.

Peter Simpson[1] Nic Waddell[2] Ana-Cristina Vargas[1] Leo Da Silva[1] Juliet French[3] Kate Hardie[3] Lynne Reid[1] Janani Jayanthan[1] and Sunil Lakhani[1]

1: Molecular & Cellular Pathology, University of Queensland Centre for Clinical Research

2: Cancer Genetics, Queensland Institute of Medical Research

3: School of Chemistry and Molecular Biosciences, University of Queensland

Archival breast tumour tissue, in the form of formalin fixed, paraffin embedded (FFPE) blocks are readily available from hospital pathology archives and provide a crucial source of research material. The degradation and cross-linking of RNA during tissue fixation and sample storage has hampered the utility of such samples. We are applying DASL (cDNA-mediated annealing, selection, extension and ligation, Illumina) technology to profile mRNA and microRNA expression patterns in breast cancer.

Breast tumour FFPE, and in some cases matched fresh frozen samples, were used. Ten micron thick tissue sections were cut onto uncoated slides and tumours were microdissected. Total RNA was extracted using a variety of methods including Trizol (Invitrogen), High Pure (Roche) and RecoverAll (Ambion). RNA quality was assessed by RPL13a Real Time RT-PCR. The DASL assay was used for molecular profiling, using the 512 Human Cancer panel, the Whole Genome DASL panel (24,000 genes) and the Human 1146 miRNA panel.

Good correlations were observed between technical replicates, patient matched FFPE - fresh frozen samples and in situ – invasive samples. Unsupervised hierarchical clustering of expression data revealed subgroups of cancers reflecting histological type and expression of ***ER, HER2, EGFR***, cytokeratin 14 and E-cadherin. Prediction of 'Sorlie' molecular subtype based on intrinsic gene list were equivocal for matched FFPE and fresh frozen samples.

DASL technology has proven useful and reliable for analysing the gene expression profiles of archival samples. Coupled with long-term clinical outcome data this provides a powerful approach to studying clinical parameters such as prognosis and outcome.

Submitted by Peter Simpson - p.simpson@uq.edu.au

39

Characterization of the melanoma miRNAome by deep sequencing

Mitchell S Stark, Sonika Tyagi, Derek Nancarrow, Glen Boyle, Richard A. Sturm and Nicholas K Hayward

Queensland Institute of Medical Research

MicroRNAs (miRNAs) are small (18-23 nucleotide) non-coding RNAs that regulate gene expression in a sequence specific manner. Little is known about the repertoire and function of miRNAs in melanoma or the melanocytic lineage. We therefore sought to undertake a comprehensive analysis of the miRNAome in a diverse range of pigment cells include melanoblasts, melanocytes, congenital naevus, acral, mucosal, cutaneous and uveal melanoma. We sequenced 12 small RNA libraries using Illumina's Genome Analyzer II. This massive parallel sequencing approach of a diverse set of melanoma and pigment cell libraries revealed a total of 539 known mature and mature-star sequences along with the prediction of 389 candidate novel miRNAs, of which 163 were common to 2 or more libraries, with 3 present in all libraries. Using the relative proportion of the total unique read counts against total number of reads, hierarchal clustering of all novel candidates plus known miRNAs gave good separation of the different histological subtypes. Some of the novel candidate miRNAs may be specific to the melanocytic lineage and as such could be used as biomarkers which could assist in the early detection of distant metastasises by measuring the circulating levels in blood. Follow up studies of the functional roles of these pigment cell miRNAs and the identification of the targets should shed further light on the development and progression of melanoma.

Submitted by Mitchell Stark - Mitchell.Stark@qimr.edu.au

40

Identification of compartment-specific anchor genes for key structures within the developing mouse kidney and transcription factor networks

Rathi D Thiagarajan, Sean M Grimmond, Kylie Georgas, Bree Rumballe, Dave Tang, and Melissa H Little

Institute for Molecular Bioscience, University of Queensland

A global analysis of gene expression across microanatomical compartments of the mouse embryonic kidney has recently been created by a combination of laser capture microdissection or FACS and microarray expression profiling. We have reanalysed this data to select "anchor" genes with absolute compartment specificity and have validated these using in situ hybridization. We have also combined whole-transcriptome sequencing to refine the transcriptional state of anchor genes and other key markers in the mouse embryonic kidney (E15.5) to address transcriptional complexity such as exon usage and promoter regions. In total, a detailed temporo-spatial expression analysis was performed on 120 genes from 5 compartments of the developing kidney (renal vesicle, early proximal tubule, medullary interstitium, ureteric tip, and medullary collecting duct), resulting in the identification of absolute compartment-specific anchor genes unique to the transcriptional landscape of the embryonic kidney. Identification of these anchor genes provide a valuable resource for further bioinformatic analysis into transcriptional regulation of tissue-specific expression. More importantly, the compartment-specific anchor genes identified here will instruct the creation of GFP:Cre animal models that will further facilitate investigations into kidney organogenesis via the engineering of GFP:Cre cassettes driven by these gene promoters. This study represents the most comprehensive search for anchor genes performed in any organ to-date and the approach taken will act as a paradigm for other organs.

Submitted by Rathi Thiagarajan - r.thiagarajan@imb.uq.edu.au

41

Expression profiling of fetal germ cell differentiation

Jocelyn van den Bergen, Denise Miles, Andrew Sinclair, and Patrick Western

ARC Centre of Excellence in Biotechnology and Development, Murdoch Children's Research Institute, Royal Children's Hospital, Parkville, Australia

The germ cell lineage is unique in that it must become highly specialized in order to support gametogenesis but must also ensure that the genome retains the complete developmental potential (totipotency) that supports development in the following generation. This is achieved through a number of mechanisms that protect the early germ cell lineage from factors inducing somatic differentiation. Part of this process involves the retained germ line expression of key pluripotency genes that protect the extraordinary developmental potential exhibited by pluripotent cell types including embryonic stem cells, embryonic germ cells and some embryonal carcinoma cells, the stem cells of testicular tumours. However, germ cells are not intrinsically pluripotent and must differentiate along the male or female pathways, a process which requires commitment of the bi-potential primordial germ cells to the spermatogenic (male) pathway and their entry into mitotic arrest, or to the oogenic pathway (females) and entry into meiosis. This involves robust regulation of regulatory networks controlling pluripotency, cell cycle and sex specific differentiation. Our work aims to further understand the mechanisms controlling differentiation, pluripotency and cell cycle in early male and female germ cells. To this end we have conducted gene (mRNA) and miRNA expression analysis using the Agilent array platform. Germ cell only populations were purified from the somatic cells of the gonad by subjecting pooled, cleanly dissected male gonad tissue to fluorescence activated cell sorting based on germ cell specific expression of an Oct4-eGFP transgene. Matched RNA and miRNA samples were isolated from three independent biological samples of E12.5-E15.5 male and female germ cells and hybridised to expression and miRNA arrays, respectively. Analysis of this data set shows that several germ line specific pathways are strongly regulated during early spermatogenic and oogenic development and that this regulation may involve specific miRNAs. Further understanding of these processes promises to lead to a greater understanding of the molecular mechanisms underlying control of pluripotency, cell cycle and differentiation in the germ line and the initiation of germ cell derived testis tumours.

Submitted by Jocelyn van den Bergan - joce.vandenbergen@mcri.edu.au

42

Exon arrays provide insight into the function of ZRANB2

Paula M. Vaz, Richard P. Grant, Fionna E. Loughlin, and Joel P. Mackay

University of Sydney

The chief processing stages of pre-mRNA, capping, splicing and polyadenylation, are thought to be closely connected with transcription of DNA. Both alternative splicing, and variation of utilised polyadenylation site and length, are avenues to regulate the content, destination and stability of the resultant mature mRNA transcript.

The SR (Ser-Arg rich) proteins are a family of splicing factors, able to bind ssRNA via one or two RNA recognition motifs and mediate protein-protein interactions using an RS domain to affect splicing.

ZRANB2 is an SR-like protein containing a C-terminal RS domain and two RanBP2-type zinc fingers at the N-terminus. We have used SELEX to show that ZRANB2 can bind ssRNA-containing tandem and possibly overlapping AGGUAA motifs and have determined the molecular basis for this binding activity by X-ray crystallography.

Recently we have used Affymetrix GeneChip Human Exon 1.0 ST Arrays to identify splicing events altered by knocking down ZRANB2 with siRNA. There are comparatively few gene-level changes, but nearly three thousand individual exons decrease by at least five fold. More importantly, of the two hundred probes with the greatest reduction, 82% of these were in the 3' untranslated region (UTR), either on or near a polyadenylation signal. We are currently validating these results using real-time PCR, and assessing the likely implications of the data for the function of ZRANB2 in the control of mRNA processing.

Submitted by Paula Vaz - pvaz2883@uni.sydney.edu.au

43

Investigation of copy number variants and familial breast cancer risk, and their interaction with rare BRCA1/2 sequence variants

Logan C Walker[1], kConFab[2], and Amanda B Spurdle[1]

1: Queensland Institute of Medical Research, Australia

2: Peter MacCallum Cancer Centre, Melbourne, Australia

A significant proportion of breast cancers arise in a subset of women who inherit genetic changes that increase risk of developing the disease. In approximately one third of these cases the increased risk of breast cancer is due inherited mutations in BRCA1 and BRCA2. However, for most women the genetic changes underlying their disease remain undetermined. Variation in breast cancer risk between individuals in the general population may be explained in part using the polygenic model of breast cancer, in which multiple genes containing low-penetrant mutations that confer low-risk combine to cause an increased but variable risk in the population. Studies have provided evidence that rare moderate-risk alleles may also contribute to breast cancer predisposition. Genomic DNA copy number variations (CNVs) are a major source of inherited human genetic variation and may play an important genetic role in breast cancer susceptibility. However, the importance of CNVs to breast cancer susceptibility has not been investigated to any great depth. Using expression arrays and high density SNP arrays, this study aims to a) determine if rare DNA CNVs in genes in BRCA1/2 related pathways cause increased risk of breast cancer in affected women with a family history of the disease, and b) determine whether BRCA1/2 pathway related CNVs interact with non-truncating variants in BRCA1/2. This research has potential to identify novel mutations that underlie increased risk for breast cancer. Results from analyses to date will be presented.

Submitted by Logan Walker - logan.walker@qimr.edu.au

44

Microarray analysis of a novel platinum drug suggests a potential for an effective treatment against cisplatin resistant tumours

Janice R. Aldrich-Wright, **Shaoyu Wang**, and Vincent J. Higgins

School of Biomedical and Health Sciences, University of Western Sydney, NSW, Australia

Drug resistance is a major problem in the treatment of tumours. The clinical application of the most widely used anticancer drug, Cisplatin, is narrowing due to drug resistance. To identify platinum drugs that can overcome the drug resistant problems we screened 30 platinum compounds for cytotoxicity against cancer and yeast cells. A lead compound (56MESS) was identified which displayed exceptional cytotoxicity. To further develop 56MESS towards clinical application, single channel Affymetrix cDNA microarray were used to explore its mechanisms of action in a yeast model organism. The array data showed that significant differentially expressed genes (at 2-fold) in response to 56MESS had less than 1% overlap with those responded to cisplatin, indicating that the mechanisms of action between these two drugs are most likely different. Exposure of 56MESS to a cisplatin resistant cancer cell line showed that this drug displayed a 100-fold higher cytotoxicity than cisplatin, suggesting that 56MESS may have potential as a treatment for cisplatin resistant tumours.

Submitted by Shaoyu Wang - s.wang@uws.edu.au

45

Microarray analysis of epididymal white adipose tissue in Klf3 -/- mice

Helen Williams, Merlin Crossley, and Kim Bell-Anderson

School of Molecular and Microbial Biosciences, Univeristy of Sydney, Australia

Krüppel-like factor 3 is a transcription factor that has been associated with adipose tissue development. Klf3 knockout (Klf3-/-) mice have less white adipose tissue (WAT) than their wildtype littermates while maintained on a standard chow diet. There are fewer adipocytes in Klf3 null fat pads and the adipocytes are smaller. Klf3-/- mice also show increased insulin sensitivity. In vitro studies have implicated Klf3 as an inhibitor of adipocyte differentiation. This experiment aimed to examine changes in gene expression in epididymal adipose tissue in Klf3-/- mice compared with their WT littermates.

RNA was extracted from epidiymal fat pads and underwent microarray analysis (Gene 1.0 ST array). Some genes commonly upregulated in obesity showed decreased expression in Klf3-/- mice, which is consistent with their lean phenotype. There were also changes in genes relating to adipocyte differentiation. These genes are currently being investigated in more detail. There was a trend towards decreased inflammatory gene expression in the Klf3-/- mice which, coupled with some decreases in genes associated with insulin resistance, could help to explain the observed differences in insulin sensitivity ***in vivo***.

The changes in gene expression observed thus far reflect the Klf3-/- phenotype of reduced number and size of adipocytes. This study, as well as future microarray studies on KLF3-/- preadipocytes and primary adipocytes, will help to elucidate how Klf3 plays a role in adipogenesis.

Submitted by Helen Williams - hwil4745@mail.usyd.edu.au

46

Development of SNP markers for differentiating dingoes from dogs and hybrids.

Auda Eltahla[2], Zi Cong Wu[2], Sven Warris[1], Webb Miller[3], Stephan Schuster[3], Alan Wilton[2,4]

1: Hanze University of Applied Sciences, The Netherlands

2: University of New South Wales, Australia

3: Pennsylvania State University, USA

4: Ramaciotti Centre for Gene Function Analysis, UNSW, Australia

Dingoes are an ancient dog breed that is distinct from European domestic dogs based on microsatellite and SNP typing. Dingoes are being replaced in the wild by breeding with domestic dogs to produce hybrids. Many hybrids look physically similar to pure dingoes. To develop a SNP based test that will be sensitive to detecting small amounts of dog ancestry in dingoes in the wild, we need loci that show differences between dogs and dingoes that are spread across the genome. Such markers would also be useful in detecting selection in some regions of genome in hybrids, for example are certain dingo traits advantageous for a existence in the wild maintained. We have sequence from a single 454 run on one dingo sample and compared it to the boxer reference genome. The run covers about 3 percent of the genome.

The comparison to dog reference genome has identified 50,000 potential sequence differences with 10,000 potential SNPs not associated with homoploymer regions. Capture arrays will be used to pull down the SNP containing sequences from 8 dogs and 8 dingoes. The captured DNA will be sequenced using 454 and markers that have the most power to discriminate between dingo and dog ancestry will be used to produce a panel of 1000 diagnostic SNPs (one allele fixed in dingoes and another in dogs) that would give a definitive level of dog ancestry for any sample and identify regions of dingo genes maintained in wild dogs.

Submitted by Alan Wilton - a.wilton@unsw.edu.au

47

Patterns across data sets: finding the cell origin of basal-like breast tumours

Di Wu and Gordon Smyth

Walter and Eliza Hall Institute of Medical Research, Melbourne

One of the major questions in microarray data analysis is to find hidden similarities between different cell types or tissue samples based on their gene expression profiles.

In this project, we have the gene expression profiles of breast tumours and 4 cell types of normal human mammary gland including stem cell enriched population, luminal progenitor, mature luminal and stromal cells.

We looked for similarities between the two datasets to reveal the cell origin of cancer subtypes. We developed methods which are able to robustly find patterns across experiments done on different tissue types using different microarray technologies. Signature genes were chosen for each normal cell type. Then scores were computed for each breast tumour sample to detect the expression signature of each normal cell type in that sample. This showed, surprisingly, that the most aggressive breast cancer subtype, basal-like tumours, shows the distinct signature, not of mammary stem cells, but of later luminal progenitor cells.

Gene set tests can be used to examine patterns in greater detail. Gene set tests confirmed conclusively that the luminal progenitor signature genes are more active in the basal-like than in other tumour subtypes. Significantly, it was also found that the luminal progenitor signature is active in mammary tissue from BRCA1 carriers who have not yet developed cancer.

Submitted by Di Wu - dwu@wehi.edu.au

48

Gene ontology testing for RNA-seq: accounting for selection bias

Matthew D Young, Matthew J Wakefield, Gordon K Smyth, and Alicia Oshlack

The Walter and Eliza Hall Institute, Melbourne

Next generation sequencing of RNA (RNA-seq) gives unprecedented detail about the transcriptional landscape of an organism. In order to accurately make use of the data, it is vital that analysis techniques are developed that take into account the technical features of RNA-seq output. Many of the specific properties of RNA-seq data are not present in previous technologies, such as microarrays, and so application of the same analysis methodologies, developed for these older technologies, may lead to bias in the results. We have previously shown that the power to detect differential expression between samples increases as the length of the transcript increases. If this feature of the data is not accounted for then GO analysis will be biased towards categories which contain many long transcripts. We have developed a new method for GO analysis of RNA-seq data which will not be biased by the length or overall read count of the genes in the categories. This method is able to correct the significance of GO categories and rank categories without inherent bias. We demonstrate the utility of the method using a previously published prostate cancer data set.

Submitted by Matthew Young - myoung@wehi.edu.au

49

Next generation sequencing in the undergraduate laboratory: sequence analysis of Wollemi Pine nuclear and chloroplast genomes

Abd Rahman, N.[1], Burgos Portugal, J.[1], Chia, S. Z.[1], Chiang, K. N. I.[1], Cho, Y.[1], Courtney, E. K.[1], Donnelly, A.[1], Fife, C.[1], Kazim, A. S. A. A.[1], Kidd, M. J.[1], Kornfeld, S. M.[1], Lee, C. G.[1], Lim, B. K.[1], Lim, W. F.[1], Lim, Y.-L. L.[1], Lo, A. W. C.[1], Ma, S. S. Q.[1], Macer-Wright, J. L.[1], Marczuk, J. C.[1], Marks-Bluth, J.[1], Mitric, S.[1], Mohamad Hanafiah, H. S.[1], Nawara, D. L[1], Nik Mohd Zamri, N. M. S.[1], O'Reilly, B.[1], Orr, M. K.[1], Othman, M. N.[1], Pan, A. Y.-H.[1], Quek, Y. J. C.[1], Souk, J. M.[1], Suffian, M. N. H.[1], Tai, Y. H.[1], Tang, F. S.[1], Wan Ahmad Kamil, W. N. I.[1], Wei, Z. Y.[1], Wong, M. K. K.[1], Yip, L. L.1, Young, K.-X. R.[1], Zhang, C. D.[1], Dawes, I.W.[2,3], Delaney, S.[2], Glenn, W.[2], Kornfeld, G.D.[2,3], Koval, J.[2,3], Marendy, E.[2], Wilkins, M.[2,3], and Wilton, A.N.[2,3]

1: Students in Recombinant DNA Techniques and Eukaryotic Molecular Biology - BIOC3281, UNSW

2: School of Biotechnology & Biomolecular Sciences, University of NSW, Sydney, NSW, 2053, Australia 3: Ramaciotti Centre for Gene Function Analysis, UNSW, Australia

First discovered in 1994 restricted to several small stands of less than 100 trees in the Wollemi National Park northwest of Sydney, the Wollemi Pine (Wollemia nobilis) is the sole extant member of a genus that is well represented in the fossil record but long thought to be extinct. The Wollemi Pine is a member of the Araucariaceae family (Order Pinales: the conifers) that also comprises (amongst others) the Bunya-bunya pine, Norfolk Island Pine, Hoop Pine, Klinki, and the Monkey-puzzle tree.

As part of their course, third year undergraduate students isolated DNA from leaves and prepared a total genomic library for sequencing. The preliminary analysis of the 454 data by the class showed that 0.49% of the reads could be mapped against the ***Pinus koraiensis*** (Korean Pine) chloroplast genome resulting in the assembly of approximately 24% of the chloroplast genome at an average 8-fold depth including several contigs that matched existing ***Wollemia nobilis*** chloroplast gene sequences in the public database. This will be extended by the class using less stringent match parameters to the Pinus koraiensis chloroplast sequence and extension of the Wollemia nobilis contigs without the scaffold. Data from this single run is sufficeient to establish phylogenetic relationships with other members of order Pinales. The data will also be used to start work on assembling the nuclear genome sequence. Given that the estimated haploid genome size of the Wollemi Pine is 13.7 GBp, it is envisioned that this project will benefit students from several classes and will form a core for teaching practical skills in modern genomics and bioinformatics at undergraduate level.

This work was supported by Roche, Illumina, Genesearch and the Ramaciotti Centre for Gene Function Analysis, UNSW.

50

SNP genotyping of samples from Riverine with mixed Australian indigenous and non-indigenous ancestry

Alan Wilton[1,2], Brian McEvoy[3], Peter Visscher[3], Eric Wang[4], Robert Moyzis[4], Sheila van Holst Pellekaan[1]

1: School of Biotechnology and Biomolecular Sciences, University of NSW

2: Ramaciotti Centre for Gene Function Analysis, University of NSW, Australia

3: Queensland Statistical Genetics, Queensland Institute of Medical Research, Herston, Queensland

4: Biological Chemistry School of Medicine, University of California, Irvine, CA, USA.

Comparison of Australian indigenous population has been compared to other ethnic population for mitochondrial DNA [van Holst Pellekaan et al 2006 Am J Phys Anthrop 131:282] and shown distinct Australian specific haplogroups which confirm the long period of genetic isolation of the population. However, relatively little information has been available for large numbers of nuclear loci in this population. This is in part due to the difficulty in satisfying the ethical issues in using samples for genetic testing and the general distrust of scientist because of past treatment. Dr van Holst has developed a close relationship with people of the Darling River region of NSW over the past 17 years through the project, 'Indigenous Australian families, genes, health and well-being'. She has obtained Aboriginal Health and Medical Research Council Ethics Committee approval and UNSW approval to extend this mitochondrial research to SNPs. Affymetrix SNP 6.0 chips from UC Irvine were processed for 38 samples at the Ramaciotti Centre. A preliminary comparison to 51 populations for 166K SNPs clusters these Riverine people in an Oceania group with Papuans and Melanesians. Although all maternal lineages are aboriginal, the mixture of indigenous and non-indigenous ancestry is evident in principal components. Comparison to 11 populations from HapMap3 for 576K SNPs separates these Riverine samples from others and confirms their mixed ancestry. We hope to extract some baseline data about variation in some Australian aborigines.

List of Registrants

Last Name	First Name	Title	Email Address
Anderson	Matthew	Mr	m.anderson@griffith.edu.au
Armstrong	Nicola	Dr	sonia@sydneybioinformatics.org
Ashton	Katie	Dr	katie.ashton@newcastle.edu.au
Bach	Warren	Dr	wbach@mscience.com.au
Bailey	Timothy	Dr	t.bailey@imb.uq.edu.au
Bailey	Lydia	Ms	t.bailey@imb.uq.edu.au
Baines	Katie	Dr	katherine.baines@newcastle.edu.au
Beckhouse	Anthony	Dr	a.beckhouse@griffith.edu.au
Blick	Tony J	Mr	blick_tony@yahoo.com.au
Borneman	Anthony	Dr	anthony.borneman@awri.com.au
Bouveret	Romaric	Dr	r.bouveret@victorchang.edu.au
Bowden	Nikola	Dr	nikola.bowden@newcastle.edu.au
Boxen	Kristine	Ms	k.boxen@auckland.ac.nz
Bragg	Lauren	Ms	lauren.bragg@csiro.au
Brown	Patrick	Prof	lowryc@stanford.edu
Bulyk	Martha	Assoc. Prof	mlbulyk@receptor.med.harvard.edu
Bundock	Peter	Dr	peter.bundock@scu.edu.au
Burdach	Jonathan	Mr	jburdach@gmail.com
Cairns	Murray	Dr	murray.cairns@newcastle.edu.au
Campbell	Ian	Assoc. Prof	ian.campbell@petermac.org
Catchpoole	Daniel	Dr	danielc@chw.edu.au
Cattley	Sonia	Dr	sonia@sydneybioinformatics.org
Charleston	Michael	Dr	mcharles@it.usyd.edu.au
Chetcuti	Albert	Dr	albertc1@chw.edu.au
Chiu	Joyce	Dr	j.chiu@unsw.edu.au
Cho	Vicky	Miss	vicky.cho@anu.edu.au
Choi	Jarny	Dr	jchoi@wehi.edu.au
Chung	Pann Pann	Miss	pann.chung@student.unsw.edu.au
Cloonan	Nicole	Dr	n.cloonan@uq.edu.au
Colakoglu	Feyza	Ms	cfey4142@mail.usyd.edu.au
Cowley	Mark	Dr	m.cowley@garvan.org.au
Cox	Mathew	Mr	mathew.b.cox@studentmail.newcastle.edu.au
Croce	Carlo	Prof	Sharon.Palko@osumc.edu
Crossley	Merlin	Prof	merlinc@usyd.edu.au
Crowe	Mark	Dr	mark.crowe@agrf.org.au
D'Avigdor	William	Mr	williama@centenary.usyd.edu.au
Dawes	Ian	Prof	i.dawes@unsw.edu.au
Day	Robert	Dr	robert.day@otago.ac.nz
De Graaf	Carolyn	Ms	degraaf@wehi.edu.au
Dewi	Vitri	Miss	v.dewi@student.usyd.edu.au
Dinger	Marcel	Dr	m.dinger@uq.edu.au
Doan	Tram	Ms	t.doan@victorchang.edu.au
Dutton-Regester	Ken	Mr	ken.dutton-regester@qimr.edu.au

Last Name	First Name	Title	Email Address
Eltahla	Auda	Mr	auda@student.unsw.edu.au
Evans	Tiffany	Ms	tiffany-jane.evans@newcastle.edu.au
Forster	Sam	Mr	sam.forster@med.monash.edu.au
French	Hugh	Mr	hugh.french@anu.edu.au
French	Cerissa	Ms	sarah.peaty@roche.com
Fung	David	Mr	dcyfung@unsw.edu.au
Funnell	Alister	Dr	a.funnell@usyd.edu.au
Gardner	Alison	Ms	alison.gardner@adelaide.edu.au
Gibson	Greg	Prof	g.gibson1@uq.edu.au
Glazov	Evgeny	Dr	e.glazov@uq.edu.au
Goel	Apurv	Mr	apurvgoel@gmail.com
Goldberg	Martin J	Snr Vice Pres	sviduka@mscience.com.au
Gonda	Thomas	Prof	t.gonda@uq.edu.au
Gould	Jodee A	Ms	jodee.gould@med.monash.edu.au
Gresham	David	Dr	dgresham@genomics.princeton.edu
Grimmond	Sean	Assoc. Prof	n.cloonan@uq.edu.au
Guinto	Jerick		sandra.staudt@agrf.org.au
Guo	Guoji		kkkggj@hotmail.com
Haviv	Izhak	Dr	izhak.haviv@petermac.org
Hayes	Vanessa	Dr	MVincent@ccia.unsw.edu.au
Henderson	Jennifer	Dr	jennifer.henderson@anu.edu.au
Henke	Robert	Dr	
Higgins	Vincent	Assoc. Prof	v.higgins@uws.edu.au
Ho	Joshua	Mr	joshua@it.usyd.edu.au
Holder	Rick	Mr	rholder@mscience.com.au
Holloway	Andrew	Dr	andrew.holloway@petermac.org
Hoopes	Laura	Prof	lhoopes@pomona.edu
Hu	Yifang	Miss	hu@wehi.edu.au
Humphreys	David	Dr	d.humphreys@victorchang.edu.au
Huttley	Gavin	Dr	Gavin.Huttley@anu.edu.au
Hyland	Ryan	Mr	r.hyland@uws.edu.au
Iyer	Vishy	Dr	vishy@icmb.utexas.edu
Jack	Briony	Ms	b.jack@usyd.edu.au
Jamaludin	M	Dr	jamal3024@yahoo.com
Janitz	Michal	Dr	m.janitz@unsw.edu.au
Janitz	Karolina	Dr	k.janitz@unsw.edu.au
Kaplan	Warren	Dr	w.kaplan@garvan.org.au
Kerfeld	Cheryl	Assoc. Prof	CKerfeld@lbl.gov
Kerr	Caroline A	Dr	caroline.kerr@csiro.au
Khan	Javed	Prof	smithpr@mail.nih.gov
Kornfeld	Geoff	Mr	g.kornfeld@unsw.edu.au
Kornfeld	Shan	Ms	shan@student.unsw.edu.au
Koval	Jason	Mr	j.koval@unsw.edu.au
Krishnan	Keerthana	Ms	k.krishnan@imb.uq.edu.au
Kulheim	Carsten	Dr	carsten.kulheim@anu.edu.au

Last Name	First Name	Title	Email Address
Lam	Hazel		sviduka@mscience.com.au
Lan	Ruiting	Dr	r.lan@unsw.edu.au
Laurie	Rebecca	Dr	becky.laurie@otago.ac.nz
Law	Charity	Miss	law@wehi.edu.au
Lin	Ruby	Dr	rubyl@unsw.edu.au
Lovell	David	Dr	david.lovell@csiro.au
Luo	Oscar	Mr	oscar.luo@anu.edu.au
Lyons	Victoria	Miss	v.lyons@uws.edu.au
MacKinnon	Ruth	Dr	ruth.mackinnon@svhm.org.au
Mak	Ka Sin	Miss	kmak3391@mail.usyd.edu.au
Marshall	Vikki	Ms	vikki.marshall@petermac.org
Matigian	Nicholas A		n.matigian@griffith.edu.au
McGavin	Cate	Ms	cate.mcgavin@agrf.org.au
McGrath	Annette		annette.mcgrath@agrf.org.au
McNoe	Les	Mr	les.mcnoe@otago.ac.nz
Mitchell	Andrew	Dr	Andrew.Mitchell@dpi.nsw.gov.au
Moraga-Martinez	Roger	Mr	roger.maraga@agresearch.co.nz
Morris	Christine	Dr	christine.morris@otago.ac.nz
Morton	Geoff	Mr	morton@cs.queensu.ca
Moscou	Frits	Mr	fmoscou@proteome.org.au
Moser	Ralf	Dr	rmoser@sequenom.com
Musgrove	Liz	Prof	l.musgrove@garvan.org.au
Nayler	Sam P	Mr	sam.nayler@qimr.edu.au
O'Doherty	Patrick	Mr	p.odoherty@uws.edu.au
O'Sullivan	Maree	Mrs	maree.osullivan@csiro.au
Palmer	Stephanie	Ms	stephanie.palmer@anu.edu
Park	Zaneta		zaneta.park-ng@agresearch.co.nz
Pearson	Richard	Dr	richard.pearson@usyd.edu.au
Peaty	Sarah	Ms	sarah.peaty@roche.com
Perreau	Victoria	Dr	vperreau@unimelb.edu.au
Peters	Greg	Dr	gregoryp@chw.edu.au
Phipson	Belinda	Mrs	phipson@wehi.edu.au
Pimanda	John	Dr	jpimanda@unsw.edu.au
Piraino	Barbara	Ms	barbara@unsw.edu.au
Preiss	Thomas	Assoc. Prof	t.preiss@victorchang.edu.au
Ritchie	William	Dr	w.ritchie@centenary.org.au
Robertson	Bronwyn	Dr	b.robertson@unsw.edu.au
Ross	Jason	Dr	jason.ross@csiro.au
Ruan	Xiaoan	Prof	ruanx@gis.a-star.edu.sg
Sahlgren	Michaela	Ms	thomaswhitington@gmail.com
Schuster	Stephan	Prof	scs@bx.psu.edu
Scott	Rodney	Prof	rodney.scott@newcastle.edu.au
Shackel	Nicholas	Dr	williama@centenary.usyd.edu.au
Shaw	Jan M	Ms	jan.shaw@csiro.au
Shearman	Jeremy	Mr	j.shearman@student.unsw.edu.au

Last Name	First Name	Title	Email Address
Shi	Wei	Dr	shi@wehi.edu.au
Simons	Cas	Dr	c.simons@qfab.org
Simpson	Peter	Dr	p.simpson@uq.edu.au
Sloggett	Clare	Dr	clare@intersect.org.au
Smith	Lorey K	Miss	lorey.smith@petermac.org
Smyth	Gordon	Assoc. Prof	smyth@wehi.edu.au
Souter	Petra	Dr	p.souter@aims.gov.au
Speirs	Helen	Dr	h.speirs@unsw.edu.au
Stanton	Jo-Ann	Dr	jo.stanton@anatomy.otago.ac.nz
Stark	Mitchell	Mr	mitchell.stark@qimr.edu.au
Steen	Jason	Mr	j.steen@imb.uq.edu.au
Streten-Joyce	Claire	Dr	claire.streten@cdu.edu.au
Talseth-Palmer	Bente	Dr	bente.talseth-palmer@newcastle.edu.au
Thiagarajan	Rathi	Ms	r.thiagarajan@imb.uq.edu.au
Thompson	Erik	Assoc. Prof	rik@svi.edu.au
Troyanskaya	Olga	Dr	bbrinker@princeton.edu
Van Den Bergen	Jocelyn	Ms	joce.vandenbergen@mcri.edu.au
Van Der Hoek	Mark	Mr	mark.vanderhoek@imvs.sa.gov.au
Van Oppen	Madeleine	Dr	m.vanoppen@aims.gov.au
Vasilevski	Oliver	Dr	ovasilevski@mscience.com.au
Vaz	Paula	Ms	pvaz2883@uni.sydney.edu.au
Vu	Thanh	Miss	thvu5032@uni.sydney.edu.au
Wakefield	Matthew	Dr	wakefield@wehi.edu.au
Walker	Logan	Dr	logan.walker@qimr.edu.au
Wall	Meaghan	Dr	meaghan.wall@petermac.org
Wang	Shaoyu	Mr	s.wang@uws.edu.au
Whitelaw	Nadia		nadia.whitelaw@qimr.edu.au
Whitington	Thomas	Mr	t.whitington@imb.uq.edu.au
Whittle	Belinda	Ms	belinda.whittle@anu.edu.au
Williams	Liam	Dr	lc.williams@auckland.ac.nz
Williams	Rohan	Dr	rohan.williams@anu.edu.au
Williams	Helen	Ms	hwil4745@mail.usyd.edu.au
Wilton	Alan	Dr	a.wilton@unsw.edu.au
Wong	Justin	Dr	j.wong@centenary.org.au
Wong	Hong Soo	Dr	hsoowong@netvigator.com
Wu	Di	Ms	dwu@wehi.edu.au
Wu	Ming	Dr	m.wu@uws.edu.au
Wu	Jingqin	Dr	Jingqin.Wu@newcastle.edu.au
Yang	Jean Yee Hwa	Dr	jeany@maths.usyd.edu.au
Young	Matthew	Mr	myoung@wehi.edu.au
Youngson	Neil	Dr	neil.youngson@qimr.edu.au
Yucel	Duygu	Ms	dyuc2466@usyd.edu.au

AMATA 2010 Conference

14-17 September 2010 • Hotel Grand Chancellor
Hobart, Tasmania • www.amata.org.au

Program Committee Members 2010

Ms Vikki Marshall
Peter MacCallum Cancer Centre
Vikki.Marshall@petermac.org

Dr Andrew Holloway
Peter MacCallum Cancer Centre
Andrew.Holloway@petermac.org

Dr Adele Holloway
Menzies Research Institute
A.F.Holloway@utas.edu.au

Dr Joanne Dickinson
Menzies Sciences Building
Jo.Dickinson@utas.edu.au

Assoc. Prof. Gordon Smyth
Walter and Eliza Hall Institute
smyth@wehi.edu.au

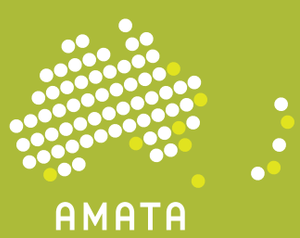
Dr Alicia Oshlack
Walter and Eiza Hall Institute
oshlack@wehi.edu.au

Dr Erik Thompson
St Vincent's Institute and University of Melbourne Department of Surgery
rik@svi.edu.au

Leishman associates
conference and event management

conference management...
it's all about people

Brigitte Cunningham
AMATA Conference Manager
Phone: 03 6234 7844
Email: brigitte@leishman-associates.com.au



www.amata.org.au