AGTA

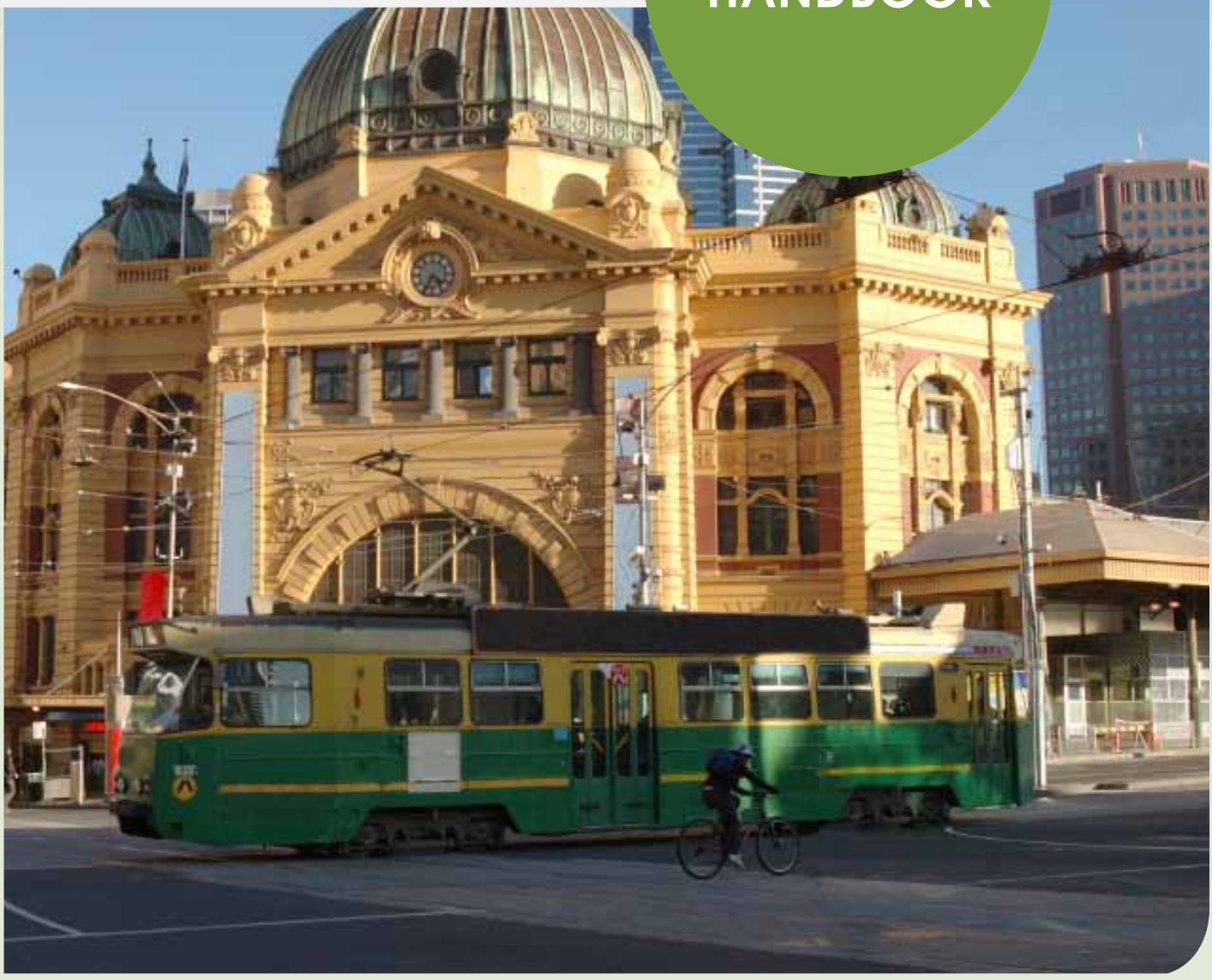
Australasian Genomic Technologies Association

2014 Conference

12–15 OCT 2014

Crown Promenade, Melbourne

HANDBOOK



SNP Genotyping Service

Custom SNP Panel Design through to Genotyping

CUSTOM PANEL DESIGN

What You Need For a Custom Panel Design

For a customised design we require customers to provide a SNP list **(rs# markers)**, as we offer an initial design consultation at **no charge**. This initial consultation allows us to work out how many multiplex assay wells are required to cover your targets of interest. From this we can then work out project pricing and provide you with an official quote.

If your project involves a species other than human or mouse, the markers will not be based upon rs# markers. In such instances, the customer will be required to submit annotated sequence information for the markers of interest.

GeneWorks

For more information, please visit the Genotyping section of the GeneWorks website.

The MassARRAY® System

The MassARRAY® system offers:

- MALDI-TOF mass spectrometry for high accuracy and sensitivity
- Robust chemistry for reproducible results
- Advanced data analysis software to meet the needs of any genomic laboratory

Agena BIOSCIENCE

How it Works

Design – The MassARRAY® Designer software enables the *in silico* design of PCR capture primers as well as iPLEX single base extension probes for multiplexed assays.

iPLEX Chemistry – The iPLEX assays incorporate a primer extension process designed to detect sequence differences at the single nucleotide level. The primer is extended, dependent upon the template sequence, resulting in an allele-specific difference in mass between extension products.

Genotype Calling –SpectroCHIP arrays spotted with samples run against multiplexed assays are placed into the MALDI-TOF mass spectrometer and the mass correlating genotype is determined in real time.

Service Pricing

As the number of multiplex wells for a panel must be known before a project can be priced, GeneWorks cannot provide a quotation based solely on the number of SNPs you are looking to genotype. Customers may provide a list of their SNPs of interest (rs# markers or annotated sequence), allowing the initial **FREE design** to be undertaken and the project scope to be defined. Official quotations include:

- Panel Design (Initial Panel Design is Free)
- Oligo manufacture and balancing of multiplexed oligo pools
- iPLEX assay
- Spotting onto MassARRAY SpectroCHIP
- Firing of SpectroCHIP on MassARRAY Analyser 4
- Analysis of genotype calls and generation of genotyping report in Excel format

Please send your SNP lists to genotyping@geneworks.com.au for a FREE initial design.

GeneWorks AUST: P 08 8159 6250 • FC 1800 882 555 E customerservice@geneworks.com.au • W www.geneworks.com.au

CONTENTS

AGTA Executive Team

Dr Ruby C Y Lin (President), Asbestos Diseases Research Institute, Sydney

Dr Carsten Kulheim, Research School of Biology, ANU, Canberra

Mr Mark van der Hoek, ACRF Cancer Genomics Facility, Adelaide

Ms Vikki Marshall, Genomics Core Facility, Murdoch Children's Research Institute, Melbourne

Dr Alicia Oshlack, Murdoch Children's Research Institute, Melbourne

Dr Richard Tothill, Research Division Cancer Therapeutics Program, Peter MacCallum Cancer Centre, Melbourne

Professor Erik (Rik) Thompson, School of Biomedical Sciences and Queensland Institute of Technology, Brisbane

Dr Jac Charlesworth, Menzies Research Institute, Hobart

Associate Professor Daniel Catchpoole, Biospecimens Research and Tumour Bank, The Children's Hospital at Westmead, Sydney

Professor Ryan Lister, The University of Western Australia, Perth

Professor Ian Paulsen, Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney

Mr Liam Williams, Centre for Genomics, Proteomics & Metabolomics, Auckland University, New Zealand

Dr Mark Waltham, St Vincent's Institute of Medical Research and Department of Surgery, University of Melbourne, Melbourne

Associate Professor Marcel Dinger, Centre for Clinical Genomics, Garvan Institute of Medical Research, Sydney

Dr Kirby Siemering, Australian Genome Research Facility (AGRF), Melbourne

Dr Mark Crowe, QFAB Bioinformatics, Brisbane

AGTA 2014 Conference Organising Committee

Dr Alicia Oshlack, Murdoch Children's Research Institute, Melbourne (Co-convenor)

Dr Richard Tothill, Peter MacCallum Cancer Centre, Melbourne (Co-convenor)

Professor Gordon Smyth, The Walter and Eliza Hall Institute of Medical Research, Melbourne

Dr Torsten Seemann, Victorian Bioinformatics Consortium, Monash University, Melbourne

Ms Vivien Vasic, MHTP Medical Genomics Facility, Monash Health Translation Precinct, Melbourne

Dr Mark Waltham, St Vincent's Institute of Medical Research, Melbourne

Dr Noel Cogan, Department of Environment and Primary Industries, Melbourne

Dr Kaylene Simpson, Peter MacCallum Cancer Centre, Melbourne

Professor Emma Whitelaw, Queensland Institute of Medical Research, Brisbane

Associate Professor Christine Wells, (AIBN), The University of Queensland, Brisbane, and Institute of Infection, Immunity and Inflammation, University of Glasgow, Scotland

Dr Andrew Fellowes, Peter MacCallum Cancer Centre, Melbourne

WELCOME FROM THE AGTA PRESIDENT

Dear Delegates and Invited Guests,

On behalf of the Australasian Genomic Technologies Association managing executives, I welcome you to the 14th Annual Scientific Meeting for AGTA in the heart of cultural Melbourne, Australia.

This meeting marks the beginning of AGTA, which transformed from AMATA to reflect the evolution of genomics in Australia. When you associate with AGTA you connect with many of Australia's leading scientists. AGTA facilitates not only rigorous discussion on genomics but innovations, biology, translational research and wet and dry laboratory techniques, across any species and various diseases. It is the premium platform where people from academia, industry, research and clinic converge to discuss major current topics impacting the 'omics' sectors. For the newcomers to the Association, this is where you form long-term friendships, network and collaborations and find job opportunities.

Our society marked a transformation in the past year with the introduction of the Small Grant Scheme that resulted in four successful recipients, announced in January 2014. Two of them will be presenting their respective posters at AGTA 2014 and two recipients will be announcing their respective bioinformatics workshops in the coming months. We face lifted our website, agtagenomics.org.au in collaboration with QFAB for hosting and ABN for Yammer, and we have a social presence on Facebook, Twitter and LinkedIn. In addition, our working groups have utilised this momentum to generate awareness of our society's benefits amongst the student populations. Furthermore, we also contributed to the McKeon Review under the leadership of ex-President A/Prof Daniel Catchpoole, and we are working towards having NHMRC and ARC to consult with us for all things 'omics".

This year's theme is aimed at enhancing our understanding on clinical sequencing with a specifically designed workshop. We have focused themes on cancer genomics, epigenomics, plant genomics, bioinformatics, transcriptomics, metagenomics and functional genomics. On behalf of the managing executives and the local conference organising committee, I would like to welcome our international plenary speakers to Melbourne and thank them for coming such a long way and taking time out of their busy schedules, and for their contributions throughout the meeting. We hope you enjoy and take time to explore our beautiful country. We are also thankful to our keynote speakers and national chairs. AGTA 2014 committee is very proud to showcase many early career researchers (ECRs) and students. For the newcomers to the Society, we trust that you will support our ECRs and students as they compete for Young Investigator Awards. The future of AGTA rests with our younger members. Australian scientists continue to capitalise on integrated technologies to tackle a diverse array of medical, agricultural and environmental, biodiversity and biosecurity questions. I would like to thank each one of you for contributing to this dynamic line-up and I hope that you enjoy the program ahead.

We are very thankful for the participation of our Gold Sponsor and Clinical Sequencing Workshop sponors Illumina, and our Gold sponsor Millennium Science incorporating Fluidigm and Pacific BioSciences. We are also thankful for our Bronze plenary session sponsors Bio-strategy, and some 19 or more exhibitors, without whom this conference would not be possible. We recognise the importance technology companies play in the evolution of genomics in Australia and you are strongly encouraged to take the time and opportunity to interact with these leading companies.

It is my pleasure to welcome you to AGTA 2014 on behalf of the AGTA managing executives and the AGTA 2014 conference organising committee. We are confident that you will enjoy AGTA 2014.

Dr Ruby CY Lin
President
Australasian Genomic Technologies Association

WELCOME FROM THE CONFERENCE CONVENORS

It is our great pleasure to welcome you to the 14th annual conference of the Australasian Genomic Technologies Association (AGTA) at the Crown Promenade Hotel, Melbourne, Australia, Sunday 12th to Wednesday 15th October 2014.

The AGTA conference (formerly known as AMATA) is Australia's foremost genomic technology conference. This year we have worked hard to bring you an exciting program that spans a diversity of research topics utilising cutting edge genomic technologies. Major themes include clinical genomics, functional genomics, bioinformatics, metagenomics, cancer genomics, epigenomics and plant genomics. We have also scheduled a workshop focusing on the application of clinical genomics, an exciting area currently undergoing a major expansion in Australia. Further features of our program include new student social functions and our regular conference dinner to be held at the Arts Centre of Melbourne under the iconic spire that has featured as part of this year's conference logo.

The AGTA conference enables the opportunity for interaction between biologists, bioinformaticians and technologists. This unique mix is one of the reasons that the Australian genomics community has a dynamic cross-disciplinary and innovative approach to genomic analysis, and is at the forefront of analysis tools for new types of 'omics data. We have invited an outstanding list of international and national speakers plus have engaged industry to showcase the best in new genomic technologies. We hope you make the most of the opportunity for some great networking and cutting edge science in the heart of the "most livable city" in the world.

We are excited to be holding the conference in Melbourne, one of the centres of Australian genomics research. We hope you enjoy the conference!

Alicia Oshlack
Murdoch Children's Research Institute, Melbourne

Richard Tothill
Peter MacCallum Cancer Centre, Melbourne

Fluidigm

THE NEW CyTOF®2 MASS CYTOMETER

Discover more. Imagine more.

Mass Cytometry. The Future of Cytometry Today

Mass cytometry resolves multiple metal probes per cell with minimal signal overlap, maximizing the information obtained from a single sample. With over 100 detector channels and over 30 isotopic probes, mass cytometry provides researchers with an unparalleled ability to phenotypically and functionally profile cells from normal and diseased states.

The DVS line of reagents, instrumentation, and data analysis tools provide all the tools necessary for novel discovery applications using mass cytometry, including:

MaxPar® line of metal-conjugated probes, panel kits and labeling reagents

State-of-the-art CyTOF 2 Mass Cytometer

High dimensional analysis tools from Fluidigm Cytobank

VISIT THE MILLENNIUM SCIENCE BOOTH TO LEARN MORE

millennium® science

www.mscience.com.au

sales@mscience.com.au • 1800 678 242

GENERAL INFORMATION

Registration Desk

The Registration Desk will be located on level one, Promenade Foyer Central. Please direct any questions you may have regarding registration, accommodation or social functions to Leishman Associates staff at this desk.

Registration Desk Opening Times

Sunday 12 October 2014
2.00pm – 6.00pm

Monday 13 October 2014
7.30am – 5.15pm

Tuesday 14 October 2014
8.00am – 5.30pm

Wednesday 15 October 2014
8.30am – 3.30pm

Accommodation

If you have any queries relating to your accommodation booking first speak to the staff at your hotel or alternatively Leishman Associates staff at the Registration Desk.

Your credit card details were supplied to the hotel you have selected, as security for your booking. If you have arrived 24 hours later than your indicated arrival day you may find that you have been charged a fee. You will be responsible for all room and incidental charges on check out and may be asked for an impression of your credit card for security against these charges. This is standard policy in many hotels.

Conference Name Badges

All delegates, speakers, sponsors and exhibitors will be provided with a name badge, which must be worn at all times within the conference venue, as it is required for access to all the conference sessions and social functions.

Conference Proceedings

PowerPoints and abstracts will be available on the AGTA website following the conclusion of the conference. Speakers will be requested to sign a release form. This is not compulsory.

Conference Wi-Fi

Wireless internet will be available throughout the conference venue for the duration of the conference. To gain access to the internet, connect to 'Crown Events and Conferences' wireless network. When you begin browsing the internet you will be prompted for the login details below. **Please note that movies, music or illicit downloads are prohibited.**

Username: AGTA1
Password: WrM20p

Join the conversation at #AGTA14.

Dress Codes

Dress throughout the day is smart casual or informal business.

Emergency Medical Care

For any medical emergency please telephone 000. The staff at your hotel will have information if you require contact details for a doctor, dentist or other health professional.

Social Program Entry

The Welcome Reception is included in the cost of each full conference registration. The Conference Dinner IS NOT included in any registration type. Social events ARE NOT included in the cost of day registrations or for accompanying partners. Places for day registrants and additional guests for these events may still be available at an additional cost. Bookings can be made at the Registration Desk subject to availability.

All delegates who are registered to attend the dinner will receive a named sticker at registration. You MUST place your sticker on a table located on poster boards next to the Registration Desk. You must allocate yourself to a table no later than 11.00 am Monday 13 October 2014.

Student Functions

All conference students and early career researchers are invited to the meet-a-mentor session and casual function on Monday 13 October, however registration is required as spaces are limited. Please forward enquiries to the Registration Desk where bookings can be made subject to availability. Further information about these events can be found on pages 20.

Proudly supported by:

COMBINE

Photographs, Videos, Recording of Sessions

Delegates are not permitted to use any type of camera or recording device at any of the sessions unless written permission has been obtained from the relevant speaker.

Speakers and Speakers' Preparation Room

All speakers should present themselves to the Speakers' Preparation Room, located in Room M1, at least 4 hours before their scheduled presentation time, to upload their presentation.

Speakers are requested to assemble in their session room 5 minutes before the commencement of their session, to meet with their session chair and to familiarise themselves with the room and the audio visual equipment.

For information on the chairperson attending your session, please see the Registration Desk.

A technician will be present in the speaker's Preparation Room during registration hours. There will be facility to test and modify your presentation as required.

Oral Presentations

Please refer to the program for the time allocated for each presentation, as these do vary. The chairperson for your session will give you a 3 minute warning, however you are asked to adhere to your time allocation so that the program remains on schedule.

Poster Presentations

Posters will be displayed in the Trade Exhibition Area for the duration of the conference. Poster Session One will be held on Monday 13 October 2014 from 1.30pm to 3.15pm. Poster Session Two will be held on Tuesday 14 October 2014 from 1.00pm to 3.00pm. Further information on the Poster Sessions can be found on pages 58.

Special Diets

All catering venues have been advised of any special diet preferences you have indicated on your registration form. Please identify yourself to venue staff as they come to serve you and they will be pleased to provide you with all pre-ordered food. For day catering, there may be a specific area for special diet food, please check with catering or conference staff.

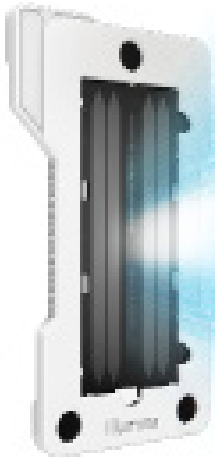
Disclaimer

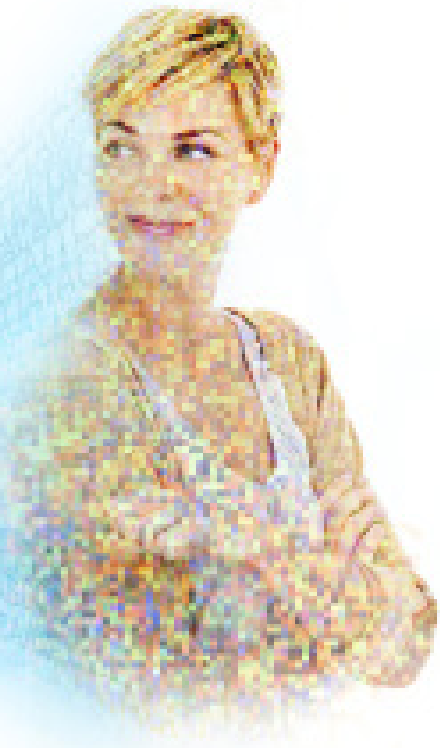
The 2014 AGTA Conference reserves the right to amend or alter any advertised details relating to dates, program and speakers if necessary, without notice, as a result of circumstances beyond their control. All attempts have been made to keep any changes to an absolute minimum.

Historical Breakthroughs

Breathtaking Progress

Radical Reinvention





Introducing **NextSeq™**
A Whole Human Genome on Your Desktop

www.illumina.com/nextseq

illumina

SUNDAY
12 October 2014

1400 – 1800	Registration Open	PROMENADE REGISTRATION DESK

OPENING ORATION
Chairs: Dr Ruby CY Lin

PROMENADE 1

1700 – 1800 GUEST SPEAKER

THE HUMAN GENOME AS THE ZIP FILE EXTRAORDINAIRE

Professor John Mattick AO FAA

Garvan Institute of Medical Research, NSW

1800 – 2000	Welcome Reception & Trade Exhibition	**Proudly sponsored by:** illumina®	PROMENADE FOYER & PROMENADE 2 & 3

DON'T FORGET to get your Exhibitor Passport stamped to go into the draw and win some great prizes – including an ipad mini!

MONDAY
13 October 2014

0730 – 1715	Registration Desk Open	PROMENADE REGISTRATION DESK
0800 – 1715	Trade Exhibition Open	PROMENADE FOYER & PROMENADE 2 & 3
0830 – 0845	Official Welcome and Conference Opening	PROMENADE 1

SESSION 1: PLANT GENOMICS
Chairs: Dr Carsten Kulheim and Dr Noel Cogan — PROMENADE 1

0845 – 0930 — KEYNOTE SPEAKER
POPULATION GENOMICS UNRAVELS GENETIC DIVERSITY AND REGULATION OF GROWTH AND DEVELOPMENT IN *EUCALYPTUS*
Professor Zander Myburg
University of Pretoria, South Africa

0930 – 1000 — INVITED SPEAKER
PLANT GENOMICS FOR CLIMATE ADAPTION
Associate Professor Justin Borevitz
Research School of Biology, ANU College of Medicine, ACT

1000 – 1015
USING SMALL RNA SEQUENCING TO CONFIRM AND DISCOVER BINDING SITES OF PENTATRICOPEPTIDE REPEAT PROTEINS
Dr Katharine Howell
ARC CoE In Plant Energy Biology, University of Western Australia, WA

1015 – 1030 — STUDENT
DISCOVERING THE MOLECULAR BASIS OF RESISTANCE IN AUSTRALIAN MYRTACEAE TO THE EXOTIC FUNGAL PATHOGEN MYRTLE RUST (PUCCINIA PSIDII
Ms Ji-Fan (Sarah) Hsieh
The Australian National University, ACT

1030 – 1045
SALT-SPECIFIC RESPONSIVE GENES ARE REVEALED BY TRANSCRIPTIONAL PROFILING OF DUNALIELLA SALINA CELLS IN RESPONSE TO A RECIPROCAL CHANGE OF SALINITY
Professor Jianhua Liu
Institute For Comprehensive Utilisation of Marine Biological Resources, Zhejiang, China

1045 – 1115 — **Morning Refreshments & Trade Exhibition** — PROMENADE FOYER & PROMENADE 2 & 3

SESSION 2: EPIGENETICS
Chairs: Professor Emma Whitelaw and Dr David Martino — PROMENADE 1

1115 – 1200 — KEYNOTE SPEAKER
3-DIMENSIONAL GENOME ORGANIZATION AND TRANSCRIPTIONAL CONTROL
Professor Melissa Fullwood
Yale-NUS College, Singapore

1200 – 1230 — INVITED SPEAKER
SHRNA SCREENS FOR NOVEL EPIGENETIC MODIFIERS AND CHARACTERISATION OF THEIR ROLES IN X INACTIVATION
Dr Marnie Blewitt
Walter and Eliza Hall Institute, VIC

1230 – 1245
CHARACTERISATION OF THE NOVEL EPIGENETIC MODIFIER RLF
Dr Harald Oey
La Trobe University, VIC

MONDAY
13 October 2014

1245 – 1300	**WIDESPREAD EPIGENOMIC AND GENDER-SPECIFIC DIFFERENCES IN ISOGENIC MICE** **Ms Helen McCormick** Victor Chang Cardiac Research Institute, NSW	
1300 – 1415	**Lunch & Trade Exhibition**	PROMENADE FOYER & PROMENADE 2 & 3
1330 – 1515	Poster Session 1	PROMENADE FOYER & PROMENADE 2 & 3
SESSION 3: BIOINFORMATICS **Chairs: Professor Gordon Smyth and Dr David Goode**		PROMENADE 1
1515 – 1600	**WHAT'S AHEAD FOR BIOLOGY? THE DATA-INTENSIVE FUTURE** **Associate Professor Titus Brown** Michigan State University, USA	
1600 – 1630	**THE STATEGRA PROJECT: NEW STATISTICAL TOOLS FOR ANALYSIS AND INTEGRATION OF DIVERSE OMICS DATA** **Dr Ana Conesa** Prince Felipe Research Center, Spain	
1630 – 1645	**CLASSIFICATION OF GENOMIC FUSIONS INTO STRUCTURAL VARIATION EVENTS** **Dr Jan Schroeder** Walter And Eliza Hall Institute, VIC	
1645 – 1700	**IDENTIFYING FUSION GENES AND OTHER TRANSCRIPTIONAL VARIATION IN CANCER THROUGH DE NOVO TRANSCRIPTOME ASSEMBLY** **Dr Nadia Davidson** Murdoch Childrens Research Institute, VIC	
1700 – 1715	**DEVELOPING AN NGS PIPELINE FOR DIAGNOSTICS** **Mr Liam McIntire** SA Pathology, SA	
1715 – 1815	**Meet-a-Mentor Session** *We invite students and early career researchers to join the conference speakers for scientific speed dating.*	
1830 onwards	**COMBINE Social Event** PJ O'Brien's, Southbank *Students and early career researchers are invited to join COMBINE for a fun night of genomics trivia.* *Walk with us from the conference venue at 18:15.*	COMBINE

TUESDAY
14 October 2014

0800 – 1730	**Registration Desk Open & Welcome Refreshments**	PROMENADE REGISTRAION DESK
0800 – 1730	Trade Exhibition Open	PROMENADE FOYER & PROMENADE 2 & 3
0825 – 0830	Welcome to Day Two	PROMENADE 1

SESSION 4: CANCER GENOMICS
Chairs: Dr Mark Waltham and Dr Ian Majewski

Proudly sponsored by: bio-strategy delivering technology

PROMENADE 1

0830 – 0915	**INTERROGATING THE ARCHITECTURE OF CANCER GENOMES** **Dr Peter Campbell** Wellcome Trust Sanger Institute, Cambridge, UK	KEYNOTE
0915 – 0945	**THE GENOMIC LANDSCAPE OF PRIMARY AND ACQUIRED RESISTANT HIGH GRADE SEROUS OVARIAN CANCER** **Professor David Bowtell** Peter MacCallum Cancer Centre, VIC	INVITED
0945 – 1015	**CIRCULATING TUMOUR DNA AS A LIQUID BIOPSY IN CANCER** **Dr Sarah-Jane Dawson** Peter MacCallum Cancer Centre, VIC	INVITED
1015 – 1030	**A NOVEL LONG NONCODING RNA, LNCUSMYCN, PROMOTES TUMOURIGENESIS BY BINDING TO THE RNA-BINDING PROTEIN NONO AND UP-REGULATING MYCN ONCOGENE EXPRESSION** **Dr Tao Liu** Histone Modification Group, Children's Cancer Institute Australia, NSW	
1030 – 1045	**PATTERNS OF CLONAL EVOLUTION INVOLVED IN TREATMENT RESISTANCE IN DIFFUSE LARGE B-CELL LYMPHOMA USING TUMOUR AND PLASMA SEQUENCING** **Dr Ryan Morin** Simon Fraser University, Canada	EARLY CAREER
1045 – 1115	**Morning Refreshments & Trade Exhibition**	PROMENADE FOYER & PROMENADE 2 & 3

SESSION 5: CLINICAL SEQUENCING
Chairs: Ms Vivien Vasic and Professor Paul Waring

PROMENADE 1

1115 – 1200	**COMPREHENSIVE GENOMIC PROFILING OF 5,000+ TUMORS REVEALS NEW INSIGHTS INTO THE DRUGGABLE GENOMIC LANDSCAPE OF SOLID TUMORS** **Dr Philip Stephens** Foundation Medicine, USA	KEYNOTE
1200 – 1230	**INTEGRATING GENOMICS INTO CLINICAL PRACTICE: A LOCAL AND INTERNATIONAL PERSPECTIVE** **Professor Kathryn North** Murdoch Childrens Research Institute, VIC	INVITED
1230 – 1300	**TRANSFERRING GENOMIC TESTING FROM RESEARCH TO DIAGNOSTICS: SIMILARITIES AND DIFFERENCES** **Professor Hamish Scott** Centre for Cancer Biology, SA	INVITED

TUESDAY
14 October 2014

1300 – 1500 Lunch, Poster session 2 & Trade Exhibition

MEETING ROOM 3 & 4

CLINICAL SEQUENCING WORKSHOP
The Clinical Sequencing Pipeline: genes to patients
Optional workshop at $50.00 per person. Bookings are essential.

1400 – 1410 INTRODUCTION AND EXISTING GUIDELINES
Dr Andrew Fellowes
Peter MacCallum Cancer Centre, VIC

1410 – 1435 NGS TECHNOLOGY: CLINICAL UTILITY AND DESIGN
Dr Desiree duSart
Victorian Clinical Genetics Services, VIC

1435 – 1500 OPTIMISATION AND VALIDATION
Dr Cliff Meldrum
Hunter New England Health, NSW

1500 – 1525 STANDARDS AND CONTROLS
Dr Natalie Thorne
Melbourne Genomics Health Alliance, VIC

1525 – 1550 Afternoon refreshments and Illumina presentation

1550 – 1615 SAMPLE PREPARATION AND SEQUENCING
Professor Graham Taylor
University of Melbourne, VIC

1615 – 1640 BIOINFORMATICS AND COMPUTING IN THE SETTING OF CLINICAL GENOMICS SETTING
Dr Mark Cowley
Garvan Institute of Medical Research, NSW

1640 – 1705 CURATION AND REPORTING OF GENETIC AND GENOMIC DATA IN THE MODERN ERA (2014!)
Professor Hamish Scott
Centre for Cancer Biology, SA

1705 – 1730 *Feedback and discussion*

PROMENADE 1

SESSION 6: TRANSCRIPTOMICS (1500 – 1730)
Chairs: Dr Nicole Cloonan and Dr Nadia Davidson

1500 – 1545 FROM ENCODE BIOCHEMICAL GENOMIC MAPS TO ENHANCER FUNCTION ASSAYS -- AND BACK: DEFINING THE ACES
Professor Barbara Wold
Caltech, USA

1545 – 1615 IDENTIFYING AND CHARACTERISING TRANSCRIPTS TARGETED BY SUBNUCLEAR-BODY ASSOCIATED PROTEINS AND LONG NONCODING RNAS
Dr Archa Fox
The Harry Perkins Institute of Medical Research, WA

TUESDAY
14 October 2014

1615 – 1645	**UNRAVELING DENDRITIC CELL-SUBSET COMMITMENT - ONE CELL AT A TIME** **Dr Andreas Schlitzer** Singapore Immunology Network, Singapore	Proudly sponsored by: millennium science | Fluidigm
1645 – 1700	**RECURSIVE SPLICING ENABLES COUPLING OF PROMOTER CHOICE TO SPLICING WITHIN LONG VERTEBRATE GENES** **Dr Christopher Sibley** University College London, England	
1700 – 1715	**INFERRING DATA-SPECIFIC MICRO-RNA FUNCTION THROUGH THE JOINT RANKING OF MICRO-RNAS AND PATHWAYS FROM SMALL SAMPLE MIRNA-SEQ AND MRNA-SEQ DATA** **Dr Ellis Patrick** University of Sydney, NSW	
1715 – 1730	**CROSS-SPECIES SYSTEMS ANALYSIS OF A METABOLIC DISEASE TO FIND A GENE SIGNATURE PREDICTIVE OF INSULIN RESISTANCE** **Dr Rima Chaudhuri** University of Sydney, NSW	
1900 – 2300	**Conference Dinner** **ANZ Pavilion, Theatres Building, Arts Centre Melbourne** *Optional function at $130.00 per person. Bookings are essential.* The Arts Centre is located on 100 St Kilda Road, Melbourne. An approximate 15 minute walk from Crown Promenade. Delegates are asked to make their own way to The Arts Centre for the Conference Dinner.	

WEDNESDAY
15 October 2014

0830 – 1530	**Registration Desk Open & Welcome Refreshments**	PROMENADE REGISTRAION DESK
0830 – 1530	Trade Exhibition Open	PROMENADE FOYER & PROMENADE 2 & 3
0855 – 0900	Welcome to Day Three	PROMENADE 1

SESSION 7: FUNCTIONAL GENOMICS
Chairs: Dr Kaylene Simpson and Dr Amee George

PROMENADE 1

0900 – 0945	**EXPLORING IMMUNE SIGNALING SYSTEMS WITH HIGH THROUGHPUT, HIGH CONTENT SCREENING** **Dr Iain Fraser** National Institutes of Health, USA	
0945 – 1015	**SYNTHETIC LETHAL TARGETING OF E-CADHERIN-DEFICIENT CANCERS** **Professor Parry Guilford** University of Otago, New Zealand	
1015 – 1030	**REPLICATION TIMING IN PROSTATE CANCER** **Dr Nicola Armstrong** University of Sydney, NSW	
1030 – 1045	**LONG NON-CODING RNA PATHOBIOLOGY IN COLORECTAL CANCER DEVELOPMENT AND PROGRESSION** **Dr Sheng Liu** Walter and Eliza Hall Institute, VIC	
1045 – 1115	**Morning Refreshments & Trade Exhibition**	PROMENADE FOYER & PROMENADE 2 & 3

SESSION 8: SINGLE MOLECULE SEQUENCING
Chairs: Dr David Lovell

PROMENADE 1

1115 – 1145	**RESOLVING GENOMIC AMBIGUITY: INSIGHTS INTO HUMAN GENOME SEQUENCING AND PATHOGEN SURVEILLANCE USING VERY LONG READS** **Dr Robert Sebra** Icahn School of Medicine, Mount Sinai Institute for Genomics and Multiscale Biology, USA **Proudly sponsored by:** millennium science | PACIFIC BIOSCIENCES	
1145 – 1300	AGTA Annual General Meeting AGTA Australasian Genomic Technologies Association agtagenomics.org.au	PROMENADE 1
1300 – 1345	**Lunch & Trade Exhibition**	PROMENADE FOYER & PROMENADE 2 & 3

KEYNOTE SPEAKER | INVITED SPEAKER | STUDENT | EARLY CAREER RESEARCHER | SPONSORED

WEDNESDAY
15 October 2014

SESSION 9: METAGENOMICS
Chairs: Dr Torsten Seemann and Dr Mark Schultz

PROMENADE 1

1345 – 1415	TOWARD ROUTINE CLINICAL SURVEILLANCE OF THE HUMAN MICROBIOTA **Associate Professor Aaron Darling** University of Technology Sydney, NSW	
1415 – 1445	THE INFANT AIRWAY MICROBIOME **Dr Kathryn Holt** Bio21 Institute, University of Melbourne, VIC	
1445 – 1500	MAPPING SHARED MARKERS ACROSS A POLYNESIAN POPULATION IN THE GENOMIC AGE **Ms Sophia Cameron-Christie** University of Otago, Otago, New Zealand	
1500 – 1515	PORE PERFORMANCE? BENCHMARKING THE OXFORD NANOPORE TECHNOLOGIES MINION FOR METAGENOMICS AND CONTIG ASSEMBLY **Dr Ken McGrath** Brisbane Node Manager, The Australian Genome Research Facility, QLD	
1515 – 1530	Awarding of Prizes 2015 Conference Launch and Conference Close	PROMENADE 1

Biology for the post-genomic era

Genome Biology

IMPACT FACTOR 10.5

Editor: Clare Garvey, PhD

- High impact – ranked 6th among research journals in Genetics and Heredity
- Open access for all research
- Fast peer review and rapid publication after acceptance
- High visibility
- Inclusion in PubMed, PubMed Central, Medline and Web of Science

For more information about the journal please visit **genomebiology.com**

biomedcentral.com

BioMed Central
The Open Access Publisher

A09435

CONFERENCE SOCIAL PROGRAM

Welcome Reception

Date: Sunday 12 October 2014
Venue: Crown Promenade, Exhibition Foyer
Time: 1800 – 2000
Dress: Business or Smart Casual

Enjoy networking with old and new acquaintances, and familiarising yourself with the trade exhibitors, whilst enjoying drinks and canapés. The Welcome Reception is included in a full registration only. Additional tickets can be purchased at $70.00 per person.

Conference Dinner

Date: Tuesday 14 October 2014
Venue: ANZ Pavilion, Theatres Building, Arts Centre Melbourne
Time: 1900 – 2300
Dress: Smart Casual

The conference dinner is the social highlight of the conference and should not be missed. Come and join us for another chance to network and meet with colleagues, whilst enjoying a great night of food, wine and dancing.

Student Functions

Meet-a-Mentor Session

Date: Monday 13 October 2014
Venue: Crown Promenade
Time: 1715 – 1815

We invite students and early career researchers to join the conference speakers for scientific speed dating. Like any good speed date, the aim will be to find out the important stuff as quickly as possible. Students and early career researches will have a short time to meet each mentor to quiz them about how to make it in science, how to have a life, and maybe even make a scientific connection. The bell will ring, and you'll get to do it all over again.

COMBINE Social Event

Date: Monday 13 October 2014
Venue: PJ O'Brien's, Southbank (10 min walk from Crown Promenade)
Time: 1830 onwards

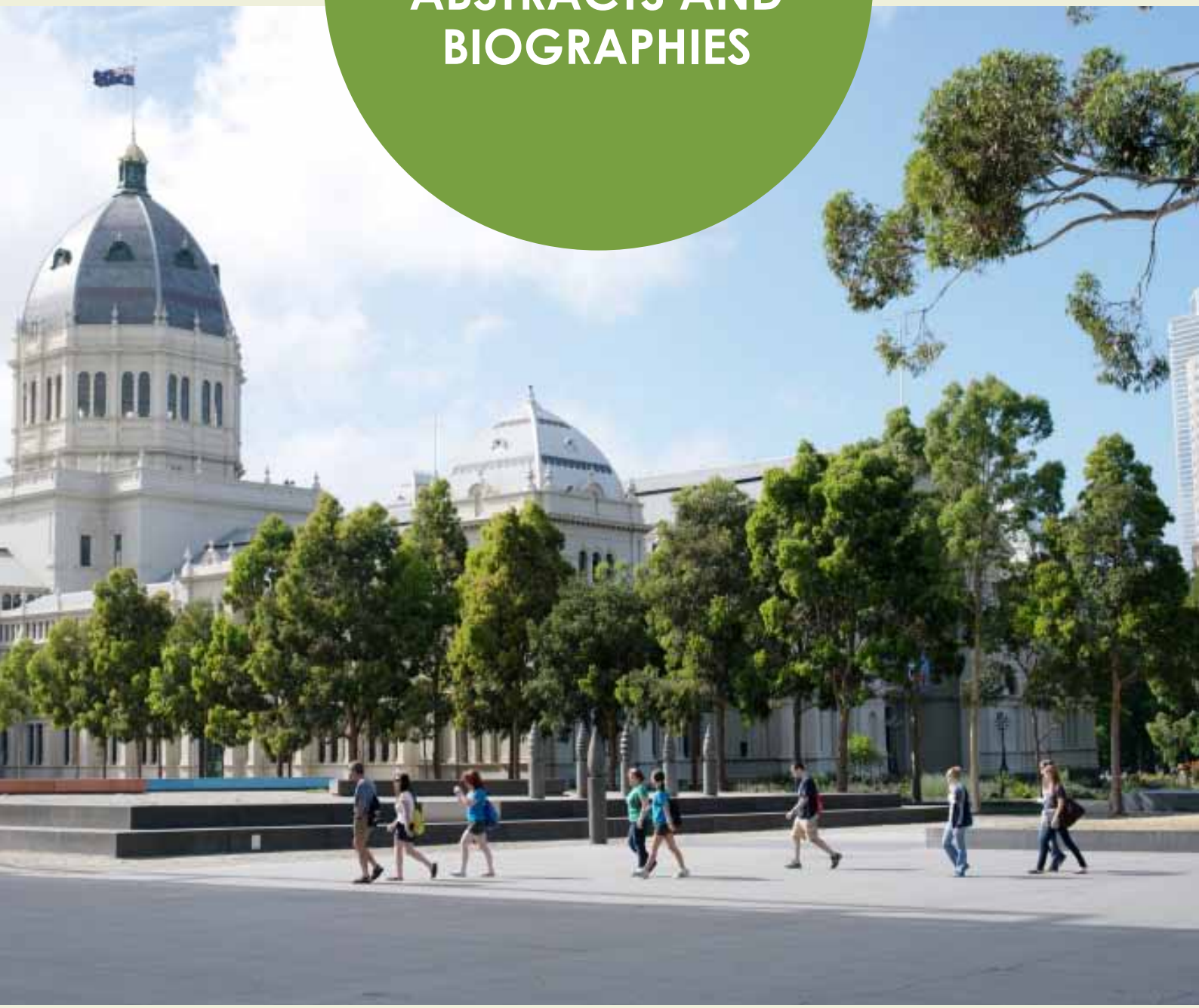
Students and early career researchers (and anyone else keen to join in!) are invited to join COMBINE for a fun night of genomics trivia. Test your knowledge of genomics technologies, bioinformatics tools, the history of sequencing and much more! As in real life genomics, your team will have a diverse range of backgrounds, so will have to work together to win the prize! Teams will form on the night, and points will be issued for both correct and creative answers.

We will provide finger food and your first few drinks. Walk with us from the conference venue at 18:15, just following the Meet-a-Mentor Session.

Proudly supported by:

COMBINE

ABSTRACTS AND BIOGRAPHIES



SUNDAY
12 October 2014

ORATION

CHAIR: DR RUBY CY LIN

1700 – 1800

THE HUMAN GENOME AS THE ZIP FILE EXTRAORDINAIRE

Professor John Mattick AO FAA
GARVAN INSTITUTE OF MEDICAL RESEARCH, NSW

BIOGRAPHY

John Mattick is the Director of the Garvan Institute of Medical Research. He undertook his undergraduate and graduate training at the University of Sydney and Monash University in Melbourne. He subsequently worked at Baylor College of Medicine in Houston, the CSIRO Division of Molecular Biology in Sydney, and the University of Queensland, where he was the Foundation Director of the Institute for Molecular Bioscience and the Australian Genome Research Facility. He has also spent research periods at the Universities of Cambridge, Oxford, Cologne and Strasbourg.

John Mattick has published over 250 papers, which have received over 25,000 citations. His honours include the Honorary Fellowship of the Royal College of Pathologists of Australasia, the inaugural Gutenberg Professorship of the University of Strasbourg, the International Union of Biochemistry and Molecular Biology Medal, Fellowship of the Australian Academy of Science, Associate Membership of the European Molecular Biology Organization, and the Human Genome Organisation's Chen Award for Distinguished Achievement in Human Genetic and Genomic Research.

ABSTRACT

High throughput analyses have shown that the vast majority of the human genome is dynamically transcribed to produce a previously hidden universe of different classes of small and large, overlapping and interlacing intronic, intergenic and antisense non-protein-coding RNAs. The transcriptome is far more complex than the genome, which is best viewed as a zip file that is unpacked in highly differentiation stage- and cell-specific patterns during development. These RNAs fulfill a wide range of regulatory functions, with miRNAs and related species being best (although not well) understood. The functions of the large/long noncoding RNAs (lncRNAs) are varied and include central roles in the formation of various differentiation-specific subnuclear organelles. However, recent evidence suggests that the major function of lncRNAs is to guide chromatin-modifying complexes to their sites of action, to specify the architectural trajectories of development. Moreover, the coding and noncoding transcriptome is far from fully characterized, with RNA CaptureSeq analyses5 revealing thousands of new exons and isoforms in human cancer gene loci, and at least 1,500 previously unknown lncRNA genes in poorly characterized 'intergenic' regions that have been associated with complex diseases. Indeed, it is emerging that variations in the sequence or expression of regulatory RNAs not only underpin phenotypic differences between individuals and species, but also play significant roles in the etiology of diseases. Moreover, the emerging transcriptomic, epigenomic and nuclear structural data point to an extraordinary precision of the 4-dimensional organization and expression of the genome that far exceeds current understanding.

MONDAY
13 October 2014

SESSION 1

PLANT GENOMICS

CHAIRS: CARSTEN KULHEIM AND NOEL COGAN

0845 – 0930

POPULATION GENOMICS UNRAVELS GENETIC DIVERSITY AND REGULATION OF GROWTH AND DEVELOPMENT IN *EUCALYPTUS*

Professor Zander Myburg
UNIVERSITY OF PRETORIA, SOUTH AFRICA

BIOGRAPHY

Zander Myburg is a professor enetics at the University of Pretoria (UP) and holds the Chair in Forest Genomics and Biotechnology at UP. His research programme in the Forestry and Agricultural Biotechnology Institute (FABI) and Genomics Research Institute (GRI) focuses on the genomics and molecular genetics of wood development in fast-growing forest trees and, in particular, the genetic regulation of cellulose biosynthesis in trees. His research group is pioneering the use of systems genetics approaches to unravel the genetic control of wood formation in Eucalyptus trees. He has also been the lead investigator of the US Department of Energy (DOE) funded Eucalyptus Genome Project. He has supervised 30 postgraduate (MSc and PhD) students and is author of 47 peer-reviewed papers and book chapters in the field of plant molecular genetics and genomics.

ABSTRACT

Genetic analysis in large structured tree populations using next-generation genomics technologies provides a powerful approach to dissect genetic variation affecting regulatory networks and biosynthetic pathways underlying wood development in Eucalyptus trees. We have performed comprehensive genome mapping, transcriptome sequencing, metabolic profiling and wood cell wall analyses in an interspecific F2 backcross population of E. grandis x E. urophylla. Transcriptome profiling (Illumina RNA-seq) of developing wood of 283 F2 progeny allowed genome-wide mapping of expression QTLs (eQTLs) for more than 20,000 xylem expressed genes. Several trans-eQTL hotspots in both backcrosses mark the location of polymorphisms affecting secondary cell wall (SCW) biosynthetic genes and provide evidence for segregating components of the transcriptional network regulating SCW formation. We further analysed the genetic architecture of transcript level variation for 353 xylem expressed transcription factors (TFs) and show that shared eQTLs and expression correlation of TF genes can be used to reconstruct modules of the SCW transcriptional network and identify candidates affecting key biological processes in xylem development. We also analysed sequence variation in the transcriptomes of the F2 progeny and identified 266,433 SNPs and 9,682 indels segregating in the backcross progeny, half of which are predicted to affect protein sequence and structure and thereby potentially contribute to phenotypic variation. A total of 1,183 predicted loss-of-function (LoF) mutations are polymorphic in the backcross and inactivate alleles of 1,285 protein-coding genes including Eucalyptus orthologs of cell wall biosynthetic genes. Together these analyses provide the foundation for gaining a systems genetics understanding of the regulation of growth and development in Eucalyptus.

0930 – 1000

PLANT GENOMICS FOR CLIMATE ADAPTION

Associate Professor Justin Borevitz
RESEARCH SCHOOL OF BIOLOGY, ANU COLLEGE OF MEDICINE, ACT

BIOGRAPHY

Justin obtained his PhD in 2002 from the University of California at San Diego with Joanne Chory studying Natural Variation in Arabidopsis light response. His postdoctoral research was with Joseph Ecker (2002-2004) at the Salk Institute focused on genomic diversity in Arabidopsis using tiling microarrays. Justin then started as assistant and associate professor in the Dept of Ecology and Evolution at the University of Chicago (2004 until 2012). His research focused on Genome Wide Association Studies in Arabidopsis and next generation genotyping by sequencing in emerging model organisms. In 2012 Justin started at the ANU where current work is identifing the genetic basis of local adaptation to seasonal climates using Phenomic and Landscape Genomic approaches in plant model organisms and foundation species.

ABSTRACT

Land use and climate change have altered agroecological environments threatening food and environmental security. Revolutions in plant genomics, phenomics and environmental control are unfolding with the promise of directly selecting regional climate ready seeds for future conditions on marginal land. I will present results harnessing natural genetic diversity in models Arabidopsis and Brachypodium along with new work in woodland foundation species of switchgrass and Eucalyptus.

1000 – 1015

USING SMALL RNA SEQUENCING TO CONFIRM AND DISCOVER BINDING SITES OF PENTATRICOPEPTIDE REPEAT PROTEINS

Dr Katharine A. Howell
UNIVERSITY OF WESTERN, WA

BIOGRAPHY

Kate Howell is a Research Assistant Professor and ARC DECRA Fellow affiliated with the ARC Centre of Excellence in Plant Energy Biology at The University of Western Australia. She received her PhD from UWA in 2006 and remained at UWA as a Research Associate until 2007. After receiving an Alexander von Humboldt Foundation Research Fellowship she moved to Germany to work at the Max Planck Institute for Molecular Plant Physiology in Potsdam. She returned to Perth and UWA in 2010 and, in 2011, received one of the inaugural ARC Discovery Early Career Researcher Awards. Her research focuses on understanding how chloroplasts (the parts of plant cells that perform photosynthesis) make RNA and protein, and communicate with the rest of the cell to enable the plant to grow and respond to stress. More recently she has been developing methods using next generation sequencing technologies to examine aspects of chloroplast gene expression.

ABSTRACT

Pentatricopeptide repeat (PPR) proteins are a large family of sequence-specific RNA binding proteins targeted to mitochondria and chloroplasts that influence the expression of organellar transcripts by altering RNA sequence, processing, turnover or translation (Barkan & Small, 2014). While they are found in all eukaryotes the PPR protein family has massively expanded in the land plant lineage. For example, in the plant model, *Arabidopsis thaliana*, the family comprises over 450 proteins. However, defining the specific RNA target(s) of PPRs has been a challenge as mutations in these proteins often result in pleiotropic effects.

Tandem repeats of PPR motifs form helices which define the RNA sequence recognised by the protein. This recognition code has recently been "cracked" (Barkan et al, 2012) allowing binding sites of PPRs to be predicted and then tested experimentally. In addition, reanalysis of data generated from small RNA sequencing experiments supports the idea that short non-coding RNA fragments that accumulate in chloroplasts are likely to be "footprints" of RNA binding proteins (Ruwe & Schmitz-Linneweber, 2012). These two recent discoveries have the potential to rapidly accelerate our ability to match PPRs with their RNA targets.

We have developed a method to analyse putative PPR "footprints" by sequencing small RNAs derived from wildtype and mutant plants. As a proof-of-concept we have been able to show for 2 PPR proteins, SOT1 and EMB2564, that putative PPR footprints are lost or reduced in the mutant plants and that these observed changes are consistent with the PPR binding sites predicted using computational methods.

1015 – 1030

DISCOVERING THE MOLECULAR BASIS OF RESISTANCE IN AUSTRALIAN MYRTACEAE TO THE EXOTIC FUNGAL PATHOGEN MYRTLE RUST (PUCCINIA PSIDII)

Ms Ji-Fan (Sarah) Hsieh

THE AUSTRALIAN NATIONAL UNIVERSITY, ACT

BIOGRAPHY

Ji-Fan (Sarah) Hsieh is a PhD student in the Research School of Biology at the Australian National University. Her research focuses on discovering of the molecular basis of resistance in Australian Myrtaceae to the exotic plant pathogen Myrtle rust, and developing molecular markers to help identifying resistant plants, providing a means of maintaining production in Myrtaceae-related rural industries (e.g. oils and culinary products such as lemon myrtle) as well as helping nurseries and contributing to restoring ecosystems.

ABSTRACT

The exotic fungal pathogen Myrtle rust (Puccinia psidii) poses a serious biosecurity threat to ecosystems and forest-based industries across the world, including Australia. More than 140 Australian native species in the family Myrtaceae are susceptible to the rust disease, including eucalypts (Eucalyptus spp.) and tea trees (Melaleuca spp.). The rust pathogen originated in South and Central America, and has spread the length of the east coast since it was discovered in Australia in April 2010. The rust affects young leaves, flowers and fruits, causing lesions on the host plant. Finding genetic markers for resistance against Myrtle rust in Myrtaceae may help identify resistant plants and provide a means of maintaining production in Myrtaceae-related rural industries (e.g. oils and culinary products such as lemon myrtle) as well as helping nurseries and contributing to restoring ecosystems.

Understanding the molecular basis of resistance and identifying potential genetic markers of pathogen resistance in important in plant breeding programs and we aim to produce molecular markers that will assist breeding of rust-resistant varieties. We sequenced the transcriptomes of resistant and susceptible individuals of Melaleuca alternifolia and M. quinquenervia before and five days post-inoculation with Myrtle rust spores. A large proportion of the assembled transcripts showed differential expression. We investigated the GO categories of the differentially expressed transcripts and found many defence-related categories up-regulated in the resistant plant. One gene family that showed marked differences between the resistant and susceptible individuals was the nucleotide-binding site leucine-rich repeat family (NBS-LRR), which is responsible for the recognition of pathogens and is the start of the signal cascade that leads to the production of defence-related proteins and plant secondary metabolites. We are currently developing molecular markers for resistance based on this gene family.

1030 – 1045

SALT-SPECIFIC RESPONSIVE GENES ARE REVEALED BY TRANSCRIPTIONAL PROFILING OF DUNALIELLA SALINA CELLS IN RESPONSE TO A RECIPROCAL CHANGE OF SALINITY

Professor Jianhua Liu

INSTITUTE FOR COMPREHENSIVE UTILISATION OF MARINE BIOLOGICAL RESOURCES, ZHEJIANG, CHINA

BIOGRAPHY

Dr Jianhua Liu is currently Professor of Ocean College, Zhejiang University, Hangzhou, China and Deputy Director of Institute for Comprehensive Utilization of Marine Biological Resources of Zhoushan, Zhoushan, China. He graduated from Fudan University, Shanghai, China and received his Ph.D. degree from Free University of Brussels, Belgium. Before his job in Zhejiang University, he was Group Leader of the Systems Biology Laboratory in the Genome Institute of Singapore. His current research efforts focus on analysis of signaling pathways and transcriptional regulatory networks that are involved in regulation of cell growth and response to environmental stress using both genetic and genomic approaches in yeast and microalgae. His laboratory also works on the marine microbiome structure, function, and diversity in the Zhoushan Archipelago area. He has published more than 40 peer-reviewed papers.

ABSTRACT

The green microalga Dunaliella salina can grow in a wide range of salinities from 0.05M to 5M of sodium chloride and has become a useful model for study of cellular adaption to salinity changes. Previous profiling studies by others showed that transcriptional changes occurred in D. salina cells upon salinity increase or decrease. However, none of the genes showed induced (or repressed) transcription upon salinity increase would be repressed (or induced) when salinity decreases. Genes displaying the same direction of transcriptional changes in response to increased or decreased salinity would not be salt-specific responsive genes. In order to search for salt-specific genes, we performed transcriptional profiling in D. salina using HiSEQ2000 technology. Based on 180 million of pair-reads (i.e., 90 base/read), we obtained a de novo transcriptome consisting of 24.5 thousand contigs (or ESTs) with the minimal FPKM of 1 that were assembled using Trinity. More than 41% (or 10 thousand) of the 24.5 thousand ESTs showed to have at least 1 Best-Hit in the non-redundant protein database (ncbi.nlm.nih.gov) using BLASTX with an cut-off e-value <1E-03. Off the 10 thousand ESTs, majority ESTs had Best-Hits originated from the sequenced green microalgal genomes. The top 2 species are Volvox carteri (e.g., 3463 Best-Hits) and Chlamydomonas reinhardtii (e.g., 2846 Best-Hit). Transcriptional profiling analysis of the reciprocal change of salinity (i.e., hyperosmotic stress: from 0.5M to 2M NaCl and hypoosmotic stress: 2M to 0.5M NaCl) indicates that numbers of differentially expressed genes trigged by hyperosmotic stress are 2-fold more than those triggered by hypoosmotic stress (i.e., 900 versus 411 genes at a cut-off of fold-change > 2 and p-value < 0.05). We find that 180 genes are differentially expressed during both hyper and hypoosmotic stresses. Correlation coefficient between transcription level changes of these 180 genes upon hyper and hypoosmotic stresses is -0.66, suggesting that these genes are potential candidates of salt-specific responsive genes. In this meeting, we will report functions of the potential salt-specific responsive genes and its implications in osmotic stress regulation in D. salina.

SESSION 2

EPIGENETICS

CHAIRS: EMMA WHITELAW AND DAVID MARTINO

1115 – 1200

3-DIMENSIONAL GENOME ORGANIZATION AND TRANSCRIPTIONAL CONTROL

Professor Melissa Fullwood
YALE-NUS COLLEGE, SINGAPORE

BIOGRAPHY

Dr Melissa J. Fullwood is a Junior Principal Investigator in the Cancer Science Institute with a joint appointment as an Assistant Professor at Yale-NUS and a Joint Principal Investigator position at the Institute of Molecular and Cell Biology, A*STAR, Singapore. Her research focuses on chromatin interactions and transcription in cancer. She completed her undergraduate degree in Biological Sciences in 2005 at Stanford University, and her PhD at the Genome Institute under the NUS Graduate School for Integrative Sciences and Engineering in 2009. She was a Lee Kuan Yew Post-Doctoral Fellow in Duke-NUS. She has 17 publications, which have been cited over 1500 times, and 2 patents. She was one of three recipients of the inaugural L'Oreal-UNESCO for Women in Science National Fellowships in Singapore in 2009, and was the international winner of the GE and Science prize in 2010. She serves on the editorial board for Scientific Reports. In 2013, she received the National Research Foundation (NRF) fellowship, which comes with a S$3.5 million grant over 5 years.

ABSTRACT

Cancer involves changes in gene expression through a variety of mechanisms, including translocations and epigenetic alterations. Recent genome-wide approaches have revealed a plethora of distal regulatory elements, some of which have been associated with cancer initiation and progression. Chromatin interactions, which are two or more distant genomic regions that come together in close spatial proximity in the 3-Dimensional environment of the cell nucleus, may facilitate cancer through mechanisms such as transcriptional control by connecting distal regulatory elements with target genes. In this talk, I will present Chromatin Interaction Analysis with Paired-End Tags (ChIA-PET) sequencing, and present my lab's recent advances in refining the ChIA-PET method, analysis, and application of this method to better understand transcriptional control in cancer cell lines, as well as future goals in terms of translation of our results to the clinic.

1200 – 1230

SHRNA SCREENS FOR NOVEL EPIGENETIC MODIFIERS AND CHARACTERISATION OF THEIR ROLES IN X INACTIVATION

Dr Marnie Blewitt
WALTER AND ELIZA HALL INSTITUTE, VIC

BIOGRAPHY

Marnie Blewitt performed her PhD with Emma Whitelaw developing a mutagenesis screen for epigenetic modifiers in the mouse, graduating in 2005. Marnie worked with Douglas Hilton at The Walter and Eliza Hall Institute for her post-doctoral studies, where she identified the role ofthe novel protein Smchd1 in X inactivation, and studied the function of polycomb group proteins in hematopoietic stem cells. This earned her the AAS Gani medal and the L'Oreal Australia Women in Science fellowship 2009. In 2010, Marnie established her own group at The Walter and Eliza Hall Institute as an ARC QEII fellow, to study the molecular mechanisms of epigenetic control.

ABSTRACT
Epigenetic marks are frequently reported to correlate with gene silencing, yet there are few circumstances when the mechanistic role of specific epigenetic marks in the process of inactivation have been characterised. X chromosome inactivation (XCI), the dosage compensation mechanism in female mammals, provides a powerful system where the initiation, establishment and maintenance of silencing can be studied for hundreds of genes in parallel. Each phase of silencing can be monitored in vitro or in vivo, as female cells transition from two active X chromosomes to one active and one inactive. While the ontogeny of X chromosome inactivation (XCI) is well described, much remains unknown about the molecular mechanisms governing this process. Therefore we are using a bespoke shRNA library to screen for epigenetic regulators of XCI, as in order to understand more about epigenetic silencing, we must first identify all the key players.

We are using our targeted library in multiple XCI screens: to identify epigenetic regulators involved in initiating, establishing or maintaining XCI. For high throughput screens, we need a rapid system to survey XCI. We have produced 2 knockin alleles of the X-linked house-keeping gene Hprt: one which produces GFP and the other mCherry under the control of the Hprt promoter. We have used these alleles and identified the H3K9 methyltransferase Setdb1 as a new regulator of the maintenance of X inactivation. Using a combination of allele-specific RNA-seq, ChIP-seq and RRBS we have placed Setdb1 within the epigenetic hierarchy of XCI, which has revealed the broader significance of H3K9 methylation in epigenetic silencing.

1230 – 1245

CHARACTERISATION OF THE NOVEL EPIGENETIC MODIFIER RLF

Dr Harald Oey
LA TROBE UNIVERSITY, VIC

BIOGRAPHY
Harald completed his university studies at Griffith University in 2007. he then worked for Professor John Mattick at the IMB in Brisbane for two years before taking on a postdoctoral research position with Professor Emma Whitelaw, first at the QIMR in Brisbane and subsequently at La Trobe University in Melbourne.

ABSTRACT
For many years our lab has been running an ENU mutagenesis screen for modifiers of transgene silencing in the mouse. The screen utilizes a GFP transgene expressed in erythrocytes in a variegated manner (expressed in some erythrocytes, silent in others), and genes that enhance or supress variegation are called modifiers of murine metastable epialleles (MommeD). Recently, we reported the initial characterisation of three novel enhancers of variegation, MommeD8, MommeD28 and MommeD34, all of which have mutations in a gene called Rearranged L-myc fusion (Rlf). Interestingly, while several MommeD alleles cause reduced DNA methylation at the GFP transgene, and an increase in GFP expression, Rlf is currently the only known MommeD causing increased methylation and a concomitant decrease in GFP expression. To investigate the function of Rlf further we have carried out a range of experiments to characterise the effect of Rlf genome-wide. These include whole genome bisulfite sequencing (~30x coverage) and mRNA-seq in multiple mutant and wildtype embryonic tissues, ChIP-seq to identify Rlf DNA binding sites and co-immunoprecipitation, followed by mass spectrometry, to identify Rlf binding partners in-vivo. We find that Rlf interacts with a histone demethylase in a

complex known to regulate transcription. Furthermore, we find that embryos null for Rlf have increased DNA methylation at a large number of distinct loci throughout the genome and that most sites have chromatin marks consistent with roles in transcriptional regulation. Interestingly, despite widespread effects on the methylome, the effect on the transcriptome is surprisingly mild.

1245 – 1300

WIDESPREAD EPIGENOMIC AND GENDER-SPECIFIC DIFFERENCES IN ISOGENIC MICE

Ms Helen McCormick

VICTOR CHANG CARDIAC RESEARCH INSTITUTE, NSW

BIOGRAPHY

Helen McCormick is a PhD student from the epigenetics laboratory at the Victor Chang Cardiac Research Institute in Sydney. Her research focuses on the nature and extent of transgenerational epigenetic inheritance in mammals, as well as gender-specific methylation. Prior to this she completed her MSc at the University of Waikato in Hamilton, New Zealand on molecular ecology of wild mice.

ABSTRACT

Epigenetic variation between individuals can provide a mechanism for phenotypic variation and differences in disease risk. While a proportion of epigenetic variation can be ascribed to genetic factors, epigenetic variation can persist even in the absence of genetic differences. To investigate epigenetic variation without genetic heterogeneity, we have performed genome-wide DNA methylation analysis on genetically identical mice. We have compared sibling groups from different families and find that, while all mice display inter-individual epigenetic variation, closely related animals are more similar to each other than to those that are more distantly related. However, we find the largest differences in genome-wide methylation to be between the genders, as, even with sex chromosomes excluded from analysis, samples separate very distinctly by gender with unsupervised hierarchical clustering methods. Comparisons of autosomal methylation in 6 male and 6 female mice showed that females are hypermethylated across all chromosomes compared to males. Using a conservative statistical approach we have identified a list of 221 highly significant gender-specific differentially methylated regions (DMRs), comprised of 1117 individual differentially methylated cytosines. Analysis with gene ontological enrichment tools shows that these regions are associated with fatty acid metabolism, iron metabolism, peroxisome function and muscle differentiation, among other processes. Furthermore, genes associated with these DMRs are primarily non-tissue specific, suggesting that these differences may be established in very early development.

SESSION 3

BIOINFORMATICS

CHAIRS: GORDON SMYTH AND DAVID GOODE

1515 – 1600

WHAT'S AHEAD FOR BIOLOGY? THE DATA-INTENSIVE FUTURE

Associate Professor Titus Brown
MICHIGAN STATE UNIVERSITY, USA

BIOGRAPHY

C. Titus Brown is an assistant professor in the Department of Computer Science and Engineering and the Department of Microbiology and Molecular Genetics. He earned his PhD ('06) in developmental molecular biology from the California Institute of Technology. Brown is director of the laboratory for Genomics, Evolution, and Development (GED) at Michigan State University. He is a member of the Python Software Foundation and an active contributor to the open source software community. His research interests include computational biology, bioinformatics, open source software development, and software engineering.

ABSTRACT

The rise of -omics over the last 20 years, and the increasing need for efficient and integrative approaches to data analysis, portends some interesting and challenging changes in biology research over the next decade.

For one example, what happens when high-density multi-omics data is available for low cost for every experiment? I will discuss both my own work on efficient sequence analysis approaches as well as a larger context of some larger-scale data investigations.

1600 – 1630

THE STATEGRA PROJECT: NEW STATISTICAL TOOLS FOR ANALYSIS AND INTEGRATION OF DIVERSE OMICS DATA.

Dr Ana Conesa
PRINCE FELIPE RESEARCH CENTER, VALENCIA, SPAIN

BIOGRAPHY

Dr Conesa is head of the Genomics of Gene Expression Lab at the Prince Felipe Research Center is Valencia, Spain. She has a degree in Agricultural Engineering and a PhD in molecular microbiology from the University of Leiden of the Netherlands. From the year 2000 she has developed her scientific career further as bioinformatician where she has specialized in transcriptomics.

Her research interests split between the functional annotation of sequence data, and the development of statistical methods for the analysis of differential gene expression and data integration. She has published over 60 scientific papers and developed popular software tools like Blast2GO and Qualimap, web-based applications (Paintomics, SEA) and R packages such as maSigPro and NOISeq. Currently her research focuses on computational approaches for the functional annotation on long non-coding RNAs and the integrative modeling of functional NGS data for biomedical applications.

At present she is scientific coordinator of the EU projects STATegra in omics data integration and the Marie Curie DEANN for the creation of the NGS network between Latin-American and European labs. She is also Management Committee member of the EU COST Action SeqAhead.

ABSTRACT
Next generation sequencing has speed up genome analysis and brought omics research closer to many organisms and biological scenarios. Today an increasing number of research projects propose the combined use of different omics platforms to investigate diverse aspects of genome functioning. These proposals ideally seek to provide complementary sources of molecular information that eventually can be put together to obtain systems biology models of biological processes. Hence, it is not rare anymore to find experimental designs involving the collection of genome, transcriptome, epigenome and even metabolome data on a particular system. However, standard methodologies for the integration of diverse omics data types are not yet ready and researchers frequently face post-experiment question on how to combine data of different nature, variability, and significance into an analysis routine that sheds more light than the analysis of individual datasets separately. The STATegra project has been conceived to address these problems and provide the genomics community with user-friendly tools for the integration of different omics data types. STATegra targets several sequencing based functional genomics methods, proteomics and metabolomics. In this presentation I will describe current results of the project including the STATegraEMS, an experiment management system for storage and annotation of complex omics experiments, experimental design issues, novel integrative visualization tools, statistical approaches to integrate RNA-seq data with other omics technologies, and data mining strategies to leverage public domain datasets in the integrative effort. I will also present the STATegRa, a new Bioconductor R package for integrative omics data analysis.

1630 – 1645

CLASSIFICATION OF GENOMIC FUSIONS INTO STRUCTURAL VARIATION EVENTS

Dr Jan Schroeder

WALTER AND ELIZA HALL INSTITUTE, VIC

ABSTRACT
A precise understanding of the structural variations (SVs) between samples of DNA is important in the study of population diversity and disease causing mutations. There are numerous methods that are designed to detect SVs in DNA sequencing data. Most of these methods do not provide a higher-level interpretation of variants, instead identifying a large set of coordinate pairs that are connected via candidate fusions – so-called break points. While some SV callers include a description of underlying genomic events, they are currently limited to a set of simple variant types (such as deletions or inversions).

Here we present an algorithm to process the output of SV calling algorithms and infer the rearrangement events that explain the observed fusions. Our method creates a graph data structure from the fusion information and looks for patterns that are characteristic of more complex rearrangement types (such as balanced translocations). Our algorithm summarises fusion events that are seemingly independent in the original output into single, more complex events (for example, an intra-chromosomal translocation involves three fusions) thus reducing the output to a more intelligible set. A better-categorised output allows for better filtering of specific events that are most relevant to the experimental circumstances.

We demonstrate that our method can improve the output of various SV algorithms (such as CREST, BreakDancer, Delly, and Socrates), identifying redundant events, and reducing false positives. Augmenting SV calls in this manner simplifies exploratory SV analysis of samples in various experimental settings, such as whole genome data, exome data, in somatic or germline tissue.

1645 – 1700

IDENTIFYING FUSION GENES AND OTHER TRANSCRIPTIONAL VARIATION IN CANCER THROUGH DE NOVO TRANSCRIPTOME ASSEMBLY

Dr Nadia Davidson
MURDOCH CHILDRENS RESEARCH INSTITUTE, VIC

BIOGRAPHY

Dr Nadia Davidson is a bioinformatician who has been working within the Oshlack group at the Murdoch Childrens Research Institute since 2011. She was originally trained in Physics and Software Engineering and completed her PhD in 2011 from the University of Melbourne on the topic of Experimental Particle Physics at the Large Hadron Collider. Her research interests include methodology development and analysis of next generation RNA sequencing data. She has been involved in a diverse set of projects that include studying sex development in birds, to identifying genomic rearrangements in cancer. A common theme to her work is de novo transcriptome assembly.

ABSTRACT

Cancer is a disease in which lesions of the genome drive tumour development. One class of genomic lesion is a fusion gene - whereby two genes are joined as a result of a translocation, producing a new and potentially oncogenic gene. Many fusions are recurrent in patient tumour populations and their identification has both research and clinical benefits. RNA sequencing (RNA-Seq) is a powerful method for profiling cancer transcriptomes to identify fusions. However, many methods for identifying fusions have low precision, require tuning of parameters or are not suitable for all RNA-Seq read lengths and layouts.

Here, we present JAFFA (https://code.google.com/p/jaffa-project/), a method that overcomes these shortfalls. JAFFA is based on the idea of comparing the sequenced cancer transcriptome to a reference transcriptome, such as GENCODE. For short reads, the cancer transcriptome is built through de novo transcriptome assembly. Validation on RNA-Seq from breast-cancer cell lines and simulation, show that JAFFA outperforms competing methods.

In addition to fusions, tumour transcriptomes harbour information on other variants of potential oncogenic significance such as SNPs, indels and mis-splicing. Based on the idea of using de novo transcriptome assemblies we also present new work to identify these transcribed variants in a cancer sample. Our results are independent of either a reference genome or transcriptome, because each cancer transcriptome within a cohort is compared to all other transcriptomes. We show examples of specific genes such as tp53 that have unique and previously unobserved transcripts in some tumours.

1700 – 1815

DEVELOPING AN NGS PIPELINE FOR DIAGNOSTICS

Mr Liam McIntyre
SA PATHOLOGY, SA

BIOGRAPHY

Liam graduated with a Bachelor of Science (Biochem/Biomed) at the University of Queensland in 2007 and then completed his Masters in Bioinformatics at the University of Queensland in 2014. Currently employed by SA pathology as a genetic informatician.

ABSTRACT

Next-generation sequencing (NGS) has reached a price point where it is viable for clinical diagnostic application, however, the development of bioinformatics pipelines remains an active area of research. We have developed a read alignment and variant calling pipeline based on BWA-MEM and the GATK best practices and have adapted it to suit amplicon and enrichment based protocols for sequencing on the MiSeq. We have compared its performance against the vendor pipeline, MiSeq Reporter, by assessing recall rate and specificity using retrospective

and prospective sample analysis. We found BWA-MEM achieves higher percentage of mapped reads and coupled with GATK's haplotype caller achieves higher recall rates and accuracy, especially for indels. Use of the GVCF mode in GATK reduces the total compute time over earlier versions of the software. Further, we have developed an in-house QC pipeline to assess target region coverage. We use this pipeline to characterise coverage for a given gene panel and to define regions commonly failing coverage, which are often associated with high GC content; we further perform detailed coverage analysis on each sample. Clinically, it is important to report not only on the confidence of a given variant call, but also on the confidence of every negative call in every gene of interest. Remaining challenges for our pipeline are calling variants in genes with pseudo gene copies, especially where there is recombination between the functional and pseudo gene copies and continuing to improve our ability to call indels, especially those larger than 10bp.

TUESDAY
14 October 2014

SESSION 4

CANCER GENOMICS

CHAIRS: MARK WALTHAM AND IAN MAJEWSKI

0830 – 0915

INTERROGATING THE ARCHITECTURE OF CANCER GENOMES

Dr Peter Campbell
WELLCOME TRUST SANGER INSTITUTE, CAMBRIDGE, UK

BIOGRAPHY

Dr Peter Campbell is Head of Cancer Genetics and Genomics at the Wellcome Trust Sanger Institute, having started a Wellcome Trust Senior Clinical Fellowship in 2010. He completed specialist training in Haematology in 2002. Following this, he completed a PhD at the University of Cambridge in the molecular pathogenesis and clinical management of myeloproliferative disorders. Since 2007, Dr Campbell has been employed at the Cancer Genome Project, Wellcome Trust Sanger Institute. In 2010, he was awarded the CR-UK Future Leaders in Cancer Research Prize.

His major interest is cancer genomics, and his recent research has been concentrated on the implementation of next generation sequencing technologies for the detection of somatically acquired genetic variants in tumour samples. One major aspiration is to develop translational applications of high-throughput genomic screening for the care of patients with cancer. Further details are available at: http://www.sanger.ac.uk/research/faculty/pcampbell/

ABSTRACT
Cancer is driven by mutation. Using massively parallel sequencing technology, we can now sequence the entire genome of cancer samples, allowing the generation of comprehensive catalogues of somatic mutations of all classes. Bespoke algorithms have been developed to identify somatically acquired point mutations, copy number changes and genomic rearrangements, which require extensive validation by confirmatory testing. The findings from our first handful of genomes illustrate the potential for next-generation sequencing to provide unprecedented insights into mutational processes, cellular repair pathways and gene networks associated with cancer development. I will also review possible applications of these technologies in a diagnostic and clinical setting, and the potential routes for translation.

0915 – 0945

THE GENOMIC LANDSCAPE OF PRIMARY AND ACQUIRED RESISTANT HIGH GRADE SEROUS OVARIAN CANCER.

Professor David Bowtell
PETER MACCALLUM CANCER CENTRE, VIC

BIOGRAPHY
After 10 years as Director of Research at Peter MacCallum Cancer Centre, Melbourne, Professor Bowtell returned to full time research in early 2010. He is Head of the Cancer Genomics and Genetics Program and a Group Leader at Peter Mac.

Professor Bowtell trained as a molecular biologist and has an extensive background human cancer genomics and signal transduction. He is Principal Investigator for the Australian Ovarian Cancer Study, one of the largest population-based cohort studies of ovarian cancer in the world, involving almost 3000 women. Professor Bowtell's lab focuses on the genomic analysis of ovarian cancer and the genetic basis of treatment resistance. His work is part of the International Cancer Genome Consortium, an ambitious program to map the genomes of the 50 most common cancer types.

ABSTRACT
It has become increasingly apparent that invasive epithelial ovarian cancer is a series of entirely distinct diseases that individually bear a closer relationship to other solid cancers than each other, and that many arise from non-ovarian sites. Therefore, a comprehensive reclassification of ovarian cancer is needed, based on underlying molecular features, and the development of treatments directed towards specific subtypes.

Our work draws on patient samples accrued from the Australian Ovarian Cancer Study (AOCS), involving ~3000 women presenting for surgery for ovarian cancer or who have recurrent disease, and more recently from a rapid autopsy study (CASCADE). We have particularly focused on high grade serous ovarian cancers (HGSC), which account for ~60% of ovarian cancer deaths, and seek to understand factors that impact on the clinical outcome of HGSC patients.

The presentation will discuss molecular subtypes of HGSC and their association with drug response and survival. I will discuss our recent unpublished studies involving the whole genome sequencing of 92 patients, including paired samples collected pre-treatment and post-resistance, providing the largest whole genome study of ovarian cancer to date. I will describe novel mutations associated with acquired drug resistance.

0945 – 1015

CIRCULATING TUMOUR DNA AS A LIQUID BIOPSY IN CANCER

Dr Sarah-Jane Dawson
PETER MACCALLUM CANCER CENTRE, VIC

BIOGRAPHY

Dr Sarah-Jane Dawson is a consultant oncologist and a clinician-scientist who obtained her medical degree from the University of Melbourne, Australia. She recently completed postdoctoral studies under the direction of Prof. Caldas and Dr Rosenfeld at the CRUK Cambridge Institute, University of Cambridge, UK and is now the head of the newly formed Molecular Biomarkers and Translational Genomics Laboratory at the Peter MacCallum Cancer Centre in Melbourne. Her current research interests lie in understanding the genomic evolution of cancer and using this information to develop noninvasive molecular biomarkers for clinical application. Her recent research has centered on the application of next generation sequencing to study circulating cell free tumor DNA (ctDNA). This work has resulted in two landmark studies demonstrating the high sensitivity of ctDNA as a molecular biomarker for disease monitoring in breast cancer, and showing the novel application of serial exome sequencing of plasma to study mechanisms of resistance to cancer therapies. Together, these studies have established a paradigm for the use of ctDNA as a 'liquid biopsy' for molecular disease monitoring in solid malignancies, and as an alternative to tissue biopsies for the study of tumor evolution during the course of disease and treatment.

ABSTRACT

Cell-free circulating DNA containing tumor-specific sequences can be identified in the plasma of cancer patients. Serial analysis of ctDNA can allow the evolving genomic landscape of a cancer to be assessed, with many potential clinical applications. Analysis of ctDNA is challenging and requires highly sensitive techniques due to the small fraction of tumor specific DNA present in the circulation, however, the application of next generation sequencing technologies is now providing new opportunities to develop ctDNA as a non-invasive 'liquid biopsy' alternative to tissue biopsies for use in cancer diagnostics and management. Our recent research has revealed the high sensitivity of ctDNA as a molecular biomarker for disease monitoring in breast cancer, and demonstrated the novel application of serial plasma DNA analysis to study mechanisms of resistance to cancer therapies. I will present an overview of these recent and exciting developments as well as an update on current ongoing research.

1015 – 1030

A NOVEL LONG NONCODING RNA, LNCUSMYCN, PROMOTES TUMOURIGENESIS BY BINDING TO THE RNA-BINDING PROTEIN NONO AND UP-REGULATING MYCN ONCOGENE EXPRESSION

Dr Tao Liu
CHILDREN'S CANCER INSTITUTE AUSTRALIA, NSW

BIOGRAPHY

Dr Tao Liu obtained his PhD from The University of New South Wales in 2000. He then worked as a post-doc with Professor Samuel Breit at St. Vincent's Centre for Applied Medical Research Sydney and Professor Glenn Marshall at Children's Cancer Institute Australia. He is now Head of Histone Modification Group at Children's Cancer Institute Australia, and an Australian Research Council Future Fellow at The University of New South Wales. In the last 5 years, Dr Liu has been working on the roles of long noncoding RNAs, histone methyltransferases and histone deacetylases in modulating gene transcription, N-Myc-regulated malignant transformation, neuroblastoma initiation and progression in vitro and in vivo, as well as the anticancer efficacy of antisense oligonucleotides targeting long noncoding RNAs, histone methyltransferase inhibitors and histone deacetylase inhibitors in vitro and in vivo.

ABSTRACT
Patients with neuroblastoma, due to the amplification of a 130 kb genomic DNA region containing the MYCN oncogene, have a survival rate of less than 10%. While MYCN has been extensively studied in the last three decades, it is unknown whether other genetic elements within the 130 kb amplicon contribute to tumorigenesis. The novel long noncoding RNA up-stream of MYCN (lncUSMycN) in the 130 kb amplicon has been recently manually annotated by the Human and Vertebrate Analysis and Annotation (HAVANA) bioinformatics team, but has not been studied. Using 5'- and 3'- rapid amplification of cDNA ends PCR, we experimentally identified 140bp extra nucleotides at the 5'- end of lncUSMycN. Analysis of 341 human neuroblastoma samples showed that the lncUSMycN gene was co-amplified with the MYCN oncogene. RNA-binding protein pull-down assays and mass spectrometry analysis identified the RNA-binding protein NonO as a lncUSMycN RNA binding protein, and RNA-immunoprecipitation assays showed NonO bound to both lncUSMycN and MYCN RNAs. While knocking-down NonO and lncUSMycN with small interfering RNAs reduced MYCN RNA expression, forced over-expression of lncUSMycN resulted in MYCN RNA up-regulation by binding to NonO. In 476 human neuroblastoma samples, high levels of lncUSMycN and NonO RNA expression correlated with high levels of MYCN RNA expression and independently predicted poor patient prognoses. Moreover, treatment with antisense oligonucleotides targeting lncUSMycN in neuroblastoma-bearing mice significantly reduced MYCN expression and blocked tumor progression. Our data therefore demonstrate the important role of the novel long noncoding RNA lncUSMycN in regulating MYCN oncogene expression and neuroblastoma oncogenesis, and provide the first evidence that amplification of long noncoding RNA genes can contribute to tumorigenesis.

1030 – 1045

PATTERNS OF CLONAL EVOLUTION INVOLVED IN TREATMENT RESISTANCE IN DIFFUSE LARGE B-CELL LYMPHOMA USING TUMOUR AND PLASMA SEQUENCING

Dr Ryan Morin
SIMON FRASER UNIVERSITY, CANADA

BIOGRAPHY
Dr Morin is an Assistant Professor, co-appointed in the department of Molecular Biology and Biochemistry at Simon Fraser University and the BC Cancer Agency's Genome Sciences Centre. He earned both M.Sc. and Ph.D. degrees from the University of British Columbia (Canada) in the laboratory of Marco Marra. Dr Morin is best known for his early work using massively parallel sequencing to identify driver mutations in leukemias and lymphomas including histone modifiers such as EZH2 and MLL2. Some of his recent recent work has been published in world-class journals such as the New England Journal of Medicine, Nature Genetics, Cancer Cell and Nature. He has contributed to 42 peer-reviewed articles and 5 reviews/book chapters on various aspects of genomics and DNA sequence analysis and his published work has been collectively cited over 5500 times.

ABSTRACT
DLBCL is a common aggressive non-Hodgkin lymphoma (NHL) and although sequencing diagnostic specimens has identified many driver genes, it is unclear whether additional genes are involved in relapse or resistance to standard treatments (R-CHOP). We performed a meta-analysis of all published data interrogating the genomes of DLBCLs. To determine the extent to which clonal evolution contributes to treatment resistance, we characterized DLBCLs following relapse using exome and genome sequencing. We have also implemented approaches to detect mutant cell-free DNA in the plasma known as circulating tumour DNA to monitor tumour evolution. We analysed ctDNA using deep amplicon sequencing and targeted capture to non-invasively detect evolution of DLBCL

tumours. By comparing diagnostic tumours and ctDNA collected at relapse, we have observed examples of branched clonal evolution in which each clone appears to derive from a common progenitor clone with a restricted set of initiating mutations and a unique set of additional mutations. By comparing the prevalence of mutations in individual genes in relapse biopsies to those in prior studies of diagnostic tumours, we have identified multiple genes that are enriched for mutations following relapse. Retrospective analysis of earlier biopsies confirms the sub-clonal representation of these mutations, suggesting that treatment has selected for populations bearing specific mutations. This study demonstrates branched clonal evolution in DLBCL including acquisition of additional driver mutations during treatment that are not clonally abundant at diagnosis. We are actively investigating the modes by which these genes and mutations may contribute to tumour aggressiveness and resistance to R-CHOP.

SESSION 5

CLINICAL SEQUENCING

CHAIRS: VIVIEN VASIC AND PAUL WARING

1115 – 1200

COMPREHENSIVE GENOMIC PROFILING OF 5,000+ TUMORS REVEALS NEW INSIGHTS INTO THE DRUGGABLE GENOMIC LANDSCAPE OF SOLID TUMORS

Dr Philip Stephens
FOUNDATION MEDICINE, USA

BIOGRAPHY

Dr Stephens joined Foundation Medicine in March 2011, bringing more than a decade of experience in cancer genomics to the company. Dr Stephens is a world-renowned expert in next-generation sequencing and cancer genome analysis and has authored numerous publications in *Nature*, *Nature Genetics*, *Nature Medicine*, *Cell* and other high-profile journals.

Prior to joining Foundation Medicine, Dr Stephens held various senior research positions during his 11-year tenure with the Cancer Genome Project at the Wellcome Trust Sanger Institute under the direction of Professor Michael Stratton. During this time, Dr Stephens was a member of the team that sequenced the first two comprehensive melanoma and lung cancer genomes, and was co-lead author in the discovery of BRAF in melanoma and ERBB2 in lung cancer.

Dr Stephens received a PhD from Oxford University.

ABSTRACT

Background: Characterization of genomic changes driving an individual's disease is essential to inform rational use of targeted therapies and treatment planning for many patients with cancer. We have developed a comprehensive, pan-solid tumor next generation sequencing (NGS)-based diagnostic assay, optimized for routine FFPE specimens including fine needle biopsies/aspirates and applied it to 5,800 patients' tumors in a CLIA-certified, CAP-accredited lab.

Methods: Hybridization capture of 3,769 exons from 236 cancer-related genes and 47 introns of 19 genes commonly rearranged was applied to ≥20ng of DNA extracted from 5,800 consecutive FFPE specimens and sequenced to a unique median depth >1,000X.

Results: 95% (5,508) of specimens were successfully profiled, yielding 2,711 unique actionable alterations across 82.1% (4,520) of cases. 68.0% (3,744) of tumors harbored ≥1 actionable alteration not assayed by available tumor-type specific tests or hotspot panels. Several novel, recurrent fusions were identified including *NTRK1*, *FGFR2* and *BRAF* fusions in pan-negative lung cancer, cholangiocarcinoma, and melanoma, respectively. Pan-negative melanoma was also significantly enriched for *NF1* loss. Constitutively activating mutations in *ESR1* were identified as a novel resistance mechanism to hormonal therapy in up to 20% of patients and *ERBB2* activating mutations were enriched in 40% and 27.3% of micropapillary urothelial carcinoma and *CDH1*-mutated lobular breast cancer, respectively. Responses to targeted therapies against actionable alterations identified in patient tumors have been reported by treating physicians.

Conclusions: Comprehensive genomic profiling identified actionable alterations in the majority of tumors profiled and identified potential additional treatment options for 68% of patients targeting alterations in genes not currently assayed by available hotspot genotyping panels.

1200 – 1230

INTEGRATING GENOMICS INTO CLINICAL PRACTICE: A LOCAL AND INTERNATIONAL PERSPECTIVE

Professor Kathryn North

MURDOCH CHILDRENS RESEARCH INSTITUTE, VIC

BIOGRAPHY

Kathryn is the Director of the Murdoch Childrens Research Institute, Director of the Victoria Clinical Genetics Service and the David Danks Professor of Child Health Research at the University of Melbourne.

Kathryn trained as a paediatrician, neurologist and clinical geneticist in Sydney and Boston. Her laboratory research interests focus on gene discovery and disease mechanism in inherited muscle disorders, as well as genes which influence normal skeletal muscle function and elite athletic performance. Her clinical research focuses on clinical trials for muscular dystrophy and neurofibromatosis type 1.

Kathryn is Chair of the NHMRC Research Committee, and in 2012 was awarded the Ramaciotti Medal for Excellence in Biomedical Research and the Member of the Order of Australia (AM) for service to medicine in the field of neuromuscular and neurogenetics research. Kathryn is one of the leaders of Melbourne Genomics Health Alliance, a collaborative approach to the integration of genomic information into everyday healthcare. Melbourne Genomics is a partnership of the Royal Melbourne Hospital, Royal Children's Hospital, Murdoch Childrens Research Institute, Walter and Eliza Hall Institute, CSIRO, Australian Genome Research Facility and the University of Melbourne. In 2013, Kathryn was invited to participate in the Global Alliance for Genomics and Health, an international consortium of more than 130 institutions across 40 countries exploring the sharing of genomic and clinical data – Kathryn is Vice Chair of the Steering Committee and Co-Chair of the Clinical Working Group. The Global Alliance will play a pivotal role in how genomic data can be linked and shared on a global level, and applied to the development of targeted therapeutics.

ABSTRACT
Genomic technologies are proving transformative, but ensuring full clinical and research benefit from their application in clinical settings requires planning and collaboration. The Melbourne Genomics Health Alliance (MGHA) is a collaboration of 10 research and healthcare organisations, with the goal of integrating genomic information into everyday healthcare. MGHA are conducting a clinically-led, prospective project evaluating the feasibility of whole exome sequencing (WES) as a single, first tier assay for germline and somatic conditions. The Alliance members have developed common ethics and consent, a common clinical bioinformatics pipeline and share genomic data via a common clinical genomics data repository. Decision-making is guided by structured processes for clinician input and a Community Advisory Group. Researchers are also an integral part of advisory and working groups.

Mimicking usual clinical practice, patients with one of five diverse germline or somatic conditions are being offered whole exome sequencing in parallel to routine investigations. Sequence data is generated by multiple diagnostic laboratories, analysis is then targeted to genes known to be related to the clinical condition using a common analytic pipeline. Research results are returned to clinicians and data is available to researchers. Exome data is linked to clinical data and a consolidated electronic view provided to clinicians and researchers. The pilot phase is being evaluated to determine the barriers, feasibility, health economics and diagnostic value of genomic sequencing. This systematic approach is designed to foster incremental change and future adoption, as well as ensure future implementation delivers a viable and sustainable system across multiple organisations.

The Melbourne Genomics Health Alliance is also a member of the Global Alliance for Genomics and Health (GA4GH), an organisation of over 200 of the world's leading biomedical research institutions, healthcare providers, information technology and life science companies, funders of research, and disease and patient advocacy organizations. The Global Alliance aims to accelerate the world-wide effort to responsibly aggregate, analyse and share large amounts of genomic and clinical information to advance the understanding, diagnosis, and treatment for cancer, inherited diseases, infectious diseases, and drug responses.

1230 – 1300

TRANSFERRING GENOMIC TESTING FROM RESEARCH TO DIAGNOSTICS: SIMILARITIES AND DIFFERENCES

Professor Hamish Scott
CENTRE FOR CANCER BIOLOGY, SA

BIOGRAPHY
Professor Hamish Scott did his PhD (1992) and first post-doc at the Women's and Children's Hospital and the University of Adelaide. During these 7 years he led the discovery of genes for 3 rare human diseases. After 11 more years, with the persistence of Professor John Hopwood and others in academia and industry, this resulted in either FDA approved therapy (2003) or clinical trials of novel therapies for these diseases.

In 1995, Hamish moved to the University of Geneva Medical School in Switzerland. His focus was, and remains, the application of genetic and genomic technologies to understand diseases processes to improve diagnoses and treatment. He led international collaborations in identification of human genes causing Down syndrome and rare forms of genetic deafness and autoimmunity (e.g. arthritis and multiple sclerosis). This continues to have profound effects on our understanding of basic biology of Down syndrome, hearing and the immune system and lead to new therapeutic strategies in these and related diseases. This was also the start of his interest in cancer and leukemia as children with Down syndrome have a low incidence of solid tumours and a high incidence of leukemia. This is also when he started to work on familial predisposition to leukemia.

Hamish relocated to the Walter and Eliza Hall Institute of Medical Research (WEHI) in Melbourne in 2000 as the Inaugural Nossal Leadership Fellow. He was appointed as a National Health and Medical Research Council (NHMRC) senior research fellow in 2001. His laboratory identified 2 additional deafness genes and described the role of a gene in reprogramming the DNA of an "adult" cell from "normal" to become a gamete (sperm or oocyte). He also described a cause of familial predisposition to leukemia.

Since returning to Adelaide in January 2008 he has been Deputy and then Head of the Department of Molecular Pathology at SA Pathology, He is an inaugural member of the Centre for Cancer Biology an Affiliate Professor in both the Schools of Medicine and Molecular and Biomedical Science at the University of Adelaide and an Adjunct Professor in the School of Pharmacy and Medical Sciences in the Division of Health Sciences of the University of South Australia. He is an NHMRC principal research fellow and a Founding Fellow of the Faculty of Science (FFSc) of the Royal College of Pathologists of Australasia (RCPA). He recently led the identification of mutations in a gene in rare families and patients that predisposed them to acute myeloid leukemia (AML), infectious diseases and lymphoedema. He has helped develop and introduce new technologies and tests for improved treatment (personalized medicine) and is a Joint Director of the ACRF Cancer Genome Facility established at SA Pathology. In these roles he has been central to introducing both somatic and germline genotyping using next generation sequencing to SA Pathology at a panel, exome and whole genome level.

ABSTRACT

The gap between research genetic testing and diagnostic genetic testing is arguably closing at a rapid rate due to next generation sequencing (NGS). This talk will highlight some of the similarities and differences between the two from the perspective of a consumer, referring and treating clinicians, research and diagnostic scientists. We will discuss examples of the current implementation of various NGS panels for both germline and somatic testing at SA Pathology.

Many NGS test have technical (laboratory and bioinformatic) limitations as to what type of mutations they will actually detect, that also affect the diagnostic yield, sensitivity and specificity of the tests. For germline testing we have trialled panels, Whole exome sequencing (WES) and Whole Genome Sequencing (WGS). Across 152 germline samples in our validation study of NGS panels we determined an overall sensitivity for SNV and indel detection (380 SNPs and pathogenic variants) of 0.9, but identified gaps in coverage, often associated with GC rich exons. Genes undergoing recombination with pseudo-gene copies eluded our current analysis by NGS. After manual review in IGV, there were no false positives. NGS testing of somatic tumor samples further needs to consider DNA quality (e.g. from FFPE samples) and amount of DNA available, especially for biopsies from lung samples. Different panels with different clinical inputs have dramatically different diagnostic yields and levels of detection of variants of unknown significance which may dramatically alter follow-up testing required or the reporting of a negative result. We will discuss two cases where standard genetic testing, and exome analyses in one case, had failed to reach a diagnosis but WGS has at least provided a partial clinical solution resulting in altered patient treatments. For somatic testing, we will give research examples from CML, AML and MDS, showing that sensitivities of detection down to $1/10^4$ cells is clinically significant while a widely accepted diagnostic standard is just 1/20. Interpretation of both germline and somatic results remains key, particularly for diagnostic reporting, clinical interpretation and action.

WORKSHOP

CLINICAL SEQUENCING WORKSHOP

The Clinical Sequencing Pipeline: genes to patients

OPTIONAL WORKSHOP AT $50.00 PER PERSON. BOOKINGS ARE ESSENTIAL.

1400 – 1410

INTRODUCTION AND EXISTING GUIDELINES

Dr Andrew Fellowes
PETER MACCALLUM CANCER CENTRE, VIC

BIOGRAPHY

Dr Fellowes has worked in the field of diagnostic molecular genetics in New Zealand for over twenty years. At the Molecular Pathology Laboratory at Canterbury Health Laboratories, he was involved in the establishment of some of the first accredited molecular genetics tests in NZ and helped create one of New Zealand's largest clinical molecular genetics laboratories which he headed from 2002 until 2011. He completed a PhD in Pathology at Otago University, which contributed important insights into the structure and function of the coagulation protein fibrinogen. In the last ten years his interest has focused on the application of automation in the molecular laboratory as a driver of quality improvement through reduced wastage, variation and manual handling. He has extensive experience in robotic systems and was an early implementer of automated molecular workflow for clinical genotyping assays. His current role at the Peter MacCallum Cancer Centre in Melbourne Australia focuses on development of robust NGS-based diagnostic assays for cancer diagnosis. Dr Fellowes is a respected member of the molecular genetics community in Australasia. He was the founding chairman of the Molecular Genetics Society of Australasia in 2005, the molecular genetics special interest group of the Human Genetics Society of Australasia (HGSA). He became a Fellow of the HGSA in the discipline of Molecular Genetics in 2008 and a founding fellow of the Faculty of Science of the Royal College of Pathologists of Australasia (RCPA) in 2011.

1410 – 1435

NGS TECHNOLOGY: CLINICAL UTILITY AND DESIGN

Dr Desirée du Sart
VICTORIAN CLINICAL GENETICS SERVICES, VIC

BIOGRAPHY

Desirée du Sart: Laboratory Head of the VCGS Molecular Genetics Laboratory, one of the largest clinical laboratories for inherited disorders in Australasia. She was awarded the Victorian Premier's Commendation Prize for achievement in Medical Research in 1998 for her PhD work, and the Victoria's Young Tall Poppies award from the Australian Institute of Political Science in 1999. She actively advocates for clinical molecular genetics within Australasia through roles within the Human Genetics Society of Australasia and the RCPA Molecular QAP, and internationally through roles within the European Quality Network and the Human Variome Project.

ABSTRACT

Next Generation Sequencing [NGS] has revolutionized molecular testing in genetic disorders. It has also revealed the complexity of gene interactions and how these come together to produce a clinical phenotype observed in patients. Insight into genetic variation in the general population is key to understanding the contribution of pathogenic variants to development of the clinical phenotype observed in patients. This is the essence of establishing the clinical utility of NGS technology for use in clinical practice. This section of the workshop will discuss what is known about human genome variation, how this information allows the identification of

pathogenic or potentially pathogenic variants by NGS technology. It will also briefly discuss the various applications and designs of NGS and the best fit options for specific clinical care outcomes.

1435 – 1500

OPTIMISATION AND VALIDATION

Dr Cliff Meldrum
HUNTER NEW ENGLAND HEALTH, NSW

BIOGRAPHY

Dr Cliff Meldrum has recently been appointed as the Genomics Project Manager for NSW Health Pathology. Cliff has a Fellowship of the Human Genetics Society of Australasia (FHGSA-Molecular Genetics) and is a Founding Fellow of the RCPA faculty of Science (FFSc-RCPA). Cliff's experience in Molecular Genetics and Molecular Pathology spans over 18 years, both within the NSW Pathology Network and also the Peter MacCallum Cancer Centre in Melbourne and as an early adopter of Next Generation Sequencing technologies he has overseen their validation, accreditation and implementation into routine diagnostics. There are a number of challenges with these technologies however; there are an even greater number of potential applications likely to dramatically change how the analysis of nucleic acids is performed in diagnostic laboratories. Applications that range from targeted regions within genes to whole genomes and that encompass many of the traditional Pathology disciplines. These applications need to be explored, evaluated and where appropriate implemented into routine laboratories.

ABSTRACT

The implementation of new technologies, no matter what they are, into clinical diagnostic laboratories must be a carefully managed process with each step of the process evaluated and controlled. Next generation sequencing technologies are a challenge if for no other reason than the fact that they are not a single technology. Rather, each application is a collection of processes assembled into both lab bench work and informatics pipelines. This is further confounded by the reality that many commercial components of such pipelines have not been designed strictly with diagnostics in mind nor are they sold with that intention. This leaves the laboratory implementing such technologies in a position where they must perform relevant validation for any assay that is not supplied as a locally accredited assay for a specific application. Key points around the optimization and validation of NGS technologies in the diagnostic setting will be presented using examples where possible.

1500 – 1525

STANDARDS AND CONTROLS

Dr Natalie Thorne
MELBOURNE GENOMICS HEALTH ALLIANCE, AUSTRALIA

BIOGRAPHY

Natalie Thorne is the clinical bioinformatics and genomics project manager for the Melbourne Genomics Health Alliance. Melbourne Genomics is linking the clinical, research and teaching strengths of its seven founding members to integrate genomic medicine into everyday healthcare for the betterment of patients.

After a degree majoring in mathematics and statistics and minoring in genetics, Natalie completed her PhD at the Walter and Eliza Hall Institute's Bioinformatics division with Prof Terry Speed and Prof Gordon Smyth (2004). She worked for Cancer Research UK at the Cambridge Research Institute with Prof Simon Tavare in the computational biology group where she worked on a variety of microarray genomics applications in cancer and analysis of data from emerging genomic technologies. After five years in the UK, she returned to the Walter and Eliza Hall Institute in a senior postdoctoral position in the statistical genetics lab with Dr Bahlo. Her studies utilised high-throughput sequencing technology with a focus on cell-free DNA in plasma during

pregnancy and in urine in the kidney transplant setting.

ABSTRACT

As genomics moves towards integrating into the health care system, diagnostic grade genomic tests need to be developed. Like all diagnostics, the use of standards and controls is crucial to controlling the quality of the test and assessing its performance to make sure it is fit for clinical use. I will review what quality control is; discuss the steps in genomic testing that need to be controlled, the resources that should be used, what exists so far and what else needs to be done.

1550 – 1615

SAMPLE PREPARATION AND SEQUENCING

Professor Graham Taylor
UNIVERSITY OF MELBOURNE, VIC

BIOGRAPHY

Graham Taylor is the Herman Professor of Genomic Medicine at the University of Melbourne, Director of the Victorian Clinical Genetics Service Labroatories and director of the Australian Human Variome Project. Current interests are the development of diagnostic applications and data pipelines for Next Generation Sequencing (NGS), including the targeted re-sequencing for the diagnosis of genetic disease, the use of NGS is tumour profiling for stratified medicine, alignment-free sequence analysis and the development of secure means data sharing.

ABSTRACT

NGS requires template in the form of a library of sequence-able molecules. These are produced from purified DNA that may come from a range of sources including blood, saliva, fresh, frozen or fixed tissue. The source and quality of the DNA will influence the choice of both sequencing technology and analysis methods. I will illustrate some examples of library preparation methods from various types of tissue and discuss their effect on the sequence data

1615 – 1640

BIOINFORMATICS AND COMPUTING IN THE SETTING OF CLINICAL GENOMICS SETTING

Dr Mark Cowley
GARVAN INSTITUTE OF MEDICAL RESEARCH, NSW

BIOGRAPHY

Mark Cowley is the head of Translational Genomics, within the Kinghorn Centre for Clinical Genomics, and an early career fellow with the Cancer Institute of NSW. His research revolves around the clinical applications of next generation sequencing, for both diagnostics and discovery, utilising his expertise in bioinformatics and genomics.

ABSTRACT

In this era of clinical genome sequencing, high quality, reproducible and reliable bioinformatics is more important than ever. The challenge is to accurately and comprehensively identify genetic variants present in a patient sample, which can ultimately be used to diagnose a patient with a genetic disorder, or determine the clinically actionable variants in a tumour. This is made particularly challenging due to the structure of the genome, littered with repetitive elements and pseudogenes, the wide distribution of genetic variant sizes, from single bases all the way to aneuploidy, the wide range of DNA amount and quality, from FFPE fine needle biopsy all the way through fresh frozen tissue, and the varied scope of next generation sequencing tests, from single genes through to whole genomes which usually results in lower overall sequencing depth the more genes that are targeted.

In the context of rare Mendelian genomics, most groups are favouring either a targeted sequencing approach with large 20-300 gene targeted panels (eg cardiomyopathy), or whole exome or genome sequencing. The bioinformatics analysis begins with extensive data quality control to determine sample quality, and to ensure that enough sequencing data has been generated to meet clinical specifications. Short sequencing reads (fastq

files) are often trimmed to discard poor quality bases at the ends of the reads, and then aligned to the human reference genome. This produces SAM or BAM files which can be viewed in most modern genome browsers (eg IGV, UCSC genome browser). Short sequencing reads are usually aligned individually, leading to errors around regions of the genome that are challenging to align to (eg known insertions and deletions), which are typically cleaned up by realigning all reads around known indels. Finally single nucleotide and small insertion and deletion variants are identified through a range of algorithms that identify recurrent evidence for a variant across multiple overlapping sequencing reads. Some groups are performing copy number variant analysis on NGS data, mostly based upon the idea of read depth, however this is challenging due to variability in sequencing depth and the patchy sequencing coverage over only targeted NGS data. Variants are then extensively annotated and ultimately filtered to those that are rare in healthy individuals, known or predicted to be damaging and associated with the patient's phenotype. The final challenging task is the clinical interpretation, determining which of the one or two variants that match the patient phenotype from a list of often hundreds of variants. Due to the larger scope of these tests, variants are usually orthogonally validated, using Sanger sequencing, prior to return of results.

In the context of cancer genomics, in the diagnostic setting, typically the DNA samples are far more challenging, usually with FFPE material from either tumour blocks or biopsy, resulting in low amounts of highly fragmented DNA. This has resulted in many groups adopting an amplicon based strategy around known cancer hotspot mutations, micropanels of single genes (eg BRCA1 and BRCA2) or when feasible, across panels of up to 300 cancer genes. Similar analysis strategies exist to Mendelian genomics, though amplicon-based sequencing data has nuances which must be accounted for in the sequence alignment and variant calling stages. Typically the sequencing depth achieved is >500x, which results in very high quality variants, which in some cases have 100% sensitivity, and as such have been clinically accredited without requiring Sanger sequencing validation.

In this presentation, I will run through the bioinformatics and computing aspects of a typical clinical genomics experiment, touching on both Mendelian and cancer genomics, and some of the pitfalls.

1640 – 1705

CURATION AND REPORTING OF GENETIC AND GENOMIC DATA IN THE MODERN ERA (2014!)

Professor Hamish Scott
CENTRE FOR CANCER BIOLOGY, SA

BIOGRAPHY

Professor Hamish Scott did his PhD (1992) and first post-doc at the Women's and Children's Hospital and the University of Adelaide. During these 7 years he led the discovery of genes for 3 rare human diseases. After 11 more years, with the persistence of Professor John Hopwood and others in academia and industry, this resulted in either FDA approved therapy (2003) or clinical trials of novel therapies for these diseases.

In 1995, Hamish moved to the University of Geneva Medical School in Switzerland. His focus was, and remains, the application of genetic and genomic technologies to understand diseases processes to improve diagnoses and treatment. He led international collaborations in identification of human genes causing Down syndrome and rare forms of genetic deafness and autoimmunity (e.g. arthritis and multiple sclerosis). This continues to have profound effects on our understanding of basic biology of Down syndrome, hearing and the immune system and lead to new therapeutic strategies in these and related diseases. This was also the start of his interest in cancer and leukemia as children with Down syndrome have a low incidence of solid tumours and a high incidence of leukemia. This is also when he started to work on familial predisposition to leukemia.

Hamish relocated to the Walter and Eliza Hall Institute of Medical Research (WEHI) in Melbourne in 2000 as the Inaugural Nossal Leadership Fellow. He was appointed as a National Health and Medical Research Council (NHMRC) senior research fellow in 2001. His laboratory identified 2 additional deafness genes and described the role of a gene in reprogramming the DNA of an "adult" cell from "normal" to become a gamete (sperm or oocyte). He also described a cause of familial predisposition to leukemia.

Since returning to Adelaide in January 2008 he has been Deputy and then Head of the Department of Molecular Pathology at SA Pathology, He is an inaugural member of the Centre for Cancer Biology an Affiliate Professor in both the Schools of Medicine and Molecular and Biomedical Science at the University of Adelaide and an Adjunct Professor in the School of Pharmacy and Medical Sciences in the Division of Health Sciences of the University of South Australia. He is an NHMRC principal research fellow and a Founding Fellow of the Faculty of Science (FFSc) of the Royal College of Pathologists of Australasia (RCPA). He recently led the identification of mutations in a gene in rare families and patients that predisposed them to acute myeloid leukemia (AML), infectious diseases and lymphoedema. He has helped develop and introduce new technologies and tests for improved treatment (personalized medicine) and is a Joint Director of the ACRF Cancer Genome Facility established at SA Pathology. In these roles he has been central to introducing both somatic and germline genotyping using next generation sequencing to SA Pathology at a panel, exome and whole genome level.

ABSTRACT

Since the late 1980s, all laboratories involved in research genetic testing and diagnostic genetic testing have always had their own form of *ad hoc* proprietary databases. In the 1990s, the ever increasing numbers of pathogenic mutations and variants of unknown significance (VUS) made it clear that data sharing was required. My own observations through that time, and until today, are that many of these attempts at data sharing were actually about data control and thus failed. There was, and is, little reward for contributors of data to centralised databases. Diverse data formats of so called locus specific databases in the early years, was also problematic. This has meant that a lot of information with potential diagnostic significance has not been, and continues to not be deposited to relevant databases.

It is obvious with the increase in the number of sequence variants being described by numerous research and diagnostic studies are valuable to share. What are the barriers to share genomic information, particularly linked to phenotypic information? What are the current initiatives to help overcome past and present problems with data sharing? What is the quality of the data in the various resources available and how should this reflect on out reporting of clinical significance of sequence variants?

SESSION 6

TRANSCRIPTOMICS

CHAIRS: NICOLE CLOONAN AND NADIA DAVIDSON

1500 – 1545

FROM ENCODE BIOCHEMICAL GENOMIC MAPS TO ENHANCER FUNCTION ASSAYS -- AND BACK: DEFINING THE ACES

Professor Barbara Wold
CALTECH, USA

BIOGRAPHY

Dr Wold is the Bren Professor of molecular biology and Director of the Beckman Institute at Caltech. She began working on genome structure and gene regulation during embryo development for her Ph.D. thesis at Caltech, and developed ways to assay cis-regulatory element function during postdoctoral work at Columbia. She established joined the biology faculty at Caltech in 1981 where she and her colleagues have focused on learning the architecture and logic of gene networks that drive cell state transitions. They study skeletal muscle development, degeneration and regeneration as a favored model system. Recent work emphasizes new ways to quantitatively map the inputs and outputs of gene networks in a genome-wide manner using "next generation" ultra-high throughput DNA sequencing, and applying these methods to muscle and brain networks.

ABSTRACT

This talk will focus on relating complex biochemical signatures in the genome, including transcription factor occupancy and histone marks, RNA-Seq output, and chromatin accessibility measurements with assays for the function of genomic elements carrying these biochemical signatures. We evaluate how well different kinds evidence, individually and in combination, predict Active Chromatin Elements (ACEs) and the nature of regulation by those elements. A related challenge is to ferret out the regulatory target(s) of any given cis-acting regulatory element. Using ChIA-PET data we find that long distance physical connection topologies are often complex, with a given ACE often contacting multiple target genes and many connection passing over intervening genes. The relationship of these physical patterns to the specific enhancer activity of several hundred ACES reveals both expected and unexpected outcomes.

1545 – 1615

IDENTIFYING AND CHARACTERISING TRANSCRIPTS TARGETED BY SUBNUCLEAR-BODY ASSOCIATED PROTEINS AND LONG NONCODING RNAS

Dr Archa Fox
THE HARRY PERKINS INSTITUTE OF MEDICAL RESEARCH, UWA, WA

BIOGRAPHY

Research Fellow and Principal Investigator at the Harry Perkins Institute of Medical Research at the University of WA in Perth. She carried out her PhD with Merlin Crossley at the University of Sydney, where she investigated the interaction between zinc-finger transcription factors in hematopoiesis; graduating in 2000. She then carried out a Post-doctoral Fellowship in Dundee, UK, working with Professor Angus Lamond, investigating how nuclear organization affects gene expression. It was during this post-doctoral period that she discovered a new subnuclar body termed Œparaspeckles[1] that have formed the basis of her ongoing research program. In 2006 she returned to Australia with an NHMRC Howard Florey Fellowship to establish her laboratory in WA.

ABSTRACT

Paraspeckles are stress-induced subnuclear bodies found in mammalian cells that are formed by specific nuclear RNA binding proteins interacting with the long noncoding RNA (lncRNA) NEAT1. Once formed, paraspeckles affect gene expression, at least in part, by the sub-nuclear sequestration of specific transcription factors. NEAT1 and paraspeckles are involved in stress responses such as viral infection and cancer progression.

Here, two different types of RNA-seq experiments aimed at dissecting paraspeckle and NEAT1 structure and function will be discussed. Firstly, using Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP) to sequence RNA footprints directly crosslinked to paraspeckle proteins. This approach has revealed binding patterns for paraspeckle proteins across the 23kb NEAT1 RNA, and has also identified non-NEAT1 RNA targets in the transcriptome. Secondly, whole transcriptome analyses of CRISPR-engineered NEAT1-/- cultured cells compared to wildtype cells. These data indicate that overall NEAT1/ paraspeckles are responsible for transcriptional activation of target genes, in line with potential roles for NEAT1 in cell proliferation through directly binding to active genes.

1615 – 1645

UNRAVELING DENDRITIC CELL-SUBSET COMMITMENT - ONE CELL AT A TIME

Dr Andreas Schlitzer
SINGAPORE IMMUNOLOGY NETWORK, SINGAPORE

Sponsored by Fluidigm and Millennium Science

ABSTRACT
Dendritic cells (DCs) arise in the bone marrow (BM) from a heterogeneous pool of precursors, which gradually matures from the earliest step of the macrophage-dendritic cell precursor (MDP), via the exclusively DC committed common dendritic cell progenitor (CDP), towards migratory DC precursors

termed pre-DCs. Circulating pre-DCs transit via the blood to seed tissues where they differentiate into two major functionally specialized DC lineages, CD8□+/CD103+ DCs and CD11b+ DCs. However, where and when progenitor commitment to DC subsets is imprinted during DCpoiesis is currently unclear. Precursor populations, defined based on the expression of a limited panel of surface markers, are likely heterogeneous and part of a developmental continuum that is subjected to "graded commitment" rather than linear development. Thus, in order to gain insights into DC precursor heterogeneity, we used single cell RNA sequencing to reveal previously unappreciated DC subset specific potential within the DC precursor compartment.

Combining single cell sequencing with conventional transcriptomic analysis, CyTOF mass cytometry and intra-femoral transfer, we identified for the first time DC subset-specific precursors as well as previously unknown checkpoints for DC lineage commitment as early as the CDP stage.

We have thereby redefined and expanded the continuum of DC precursors by identifying DC subset-specific precursors, which will allow efficient targeting and manipulation of DC subsets during health and disease.

1645 – 1700

RECURSIVE SPLICING ENABLES COUPLING OF PROMOTER CHOICE TO SPLICING WITHIN LONG VERTEBRATE GENES

Dr Christopher Sibley
UNIVERSITY COLLEGE LONDON, ENGLAND

BIOGRAPHY
Christopher Sibley is presently a visiting researcher for a year, working with Dr Mike Inouye at Melbourne University Department of Pathology. He carried out his DPhil at the University of Oxford in the lab of Professor Matthew Wood, before carrying out his post-doctoral research with Professor Jernej Ule first at the UK Medical Research Councils Laboratory of Molecular Biology in Cambridge, and now at the University College London Institute of Neurology. His work involves the development and application of functional genomics approaches to study post-transcriptional RNA networks within brain.

ABSTRACT
Recent studies have shown that expression of long genes in the mammalian brain can be perturbed by regulatory factors linked to neurodevelopmental or neurodegenerative disorders1-3, suggesting unique regulatory mechanisms. Here we find that most genes that are long in all vertebrate species are specifically expressed in the brain and contain long first introns with a high incidence of cryptic promoters. Using the functional genomics methods of total RNA-seq and cross-linked immune-precipitation (CLIP) we identified newly discovered recursive splice sites (RSS) in several of the long introns within these genes. These RSS are highly conserved and are strongly bound by the spliceosome. We show that recursive splicing requires exon definition via a "recursive exon" that is located downstream of the RSS. Importantly, RSS enable an unusually discrete regulation of splicing; the recursive exon is not detectable when the dominant promoter is used, but is completely included when cryptic promoters are used. Most recursive exons contain premature stop codons, which lead

to transcript degradation via NMD. Thus, by coordinating inclusion of recursive exons with the use of upstream promoters, RSS can control the stability of resulting transcripts. Collectively, we demonstrate how exon definition and splice site competition enable recursive splicing, a new mechanism for dictating promoter-dependent splicing and transcript stability in the brain.

1700 – 1730

INFERRING DATA-SPECIFIC MICRO-RNA FUNCTION THROUGH THE JOINT RANKING OF MICRO-RNAS AND PATHWAYS FROM SMALL SAMPLE MIRNA-SEQ AND MRNA-SEQ DATA

Dr Ellis Patrick
UNIVERSITY OF SYDNEY, NSW

BIOGRAPHY

Ellis Patrick is currently a postdoctoral research associate in the School of Mathematics and Statistics at the University of Sydney. He identifies as a statistical bioinformatician and his main research interests are aligned with both the horizontal and vertical integration of large datasets.

ABSTRACT

Understanding how micro-RNA regulate messenger-RNA is a fundamental biological problem. In practice, identifying and interpreting the functional impacts of the regulatory relationships between micro-RNA and messenger-RNA is non-trivial. By leveraging the information within existing

micro-RNA target and pathway databases we propose an approach to stabilise the estimation and annotation of micro-RNA regulation that is suitable for datasets with small sample sizes. Specifically, we propose a supervised framework, built upon concepts of significance combination, for jointly ranking regulatory micro-RNA and their potential functional impacts with respect to a condition of interest.

Our proposed method, pMimCor, directly tests if a micro-RNA is differentially expressed and if its predicted targets that lie in a common biological pathway have changed in the opposite direction. Utilising a database of predicted micro-RNA regulatory targets and pathway annotations, pMimCor circumvents much of the need to produce large datasets in order to generate stable and accurate network estimation. Furthermore, by ranking relevant micro-RNA and pathways concurrently pMimCor also provides highly interpretable biological information that can be used to infer how a given micro-RNA regulates specific functional processes.

We demonstrate this framework can successfully be applied to datasets with small sample sizes, with pMimCor performing competitively across a range of performance metrics when compared to other approaches. When applied to publicly available cancer datasets and our own small Notch2 conditional knockout experiment with mice, pMimCor identifies more micro-RNA that have been associated with each specific dataset in the literature.

1715 – 1730

CROSS-SPECIES SYSTEMS ANALYSIS OF A METABOLIC DISEASE TO FIND A GENE SIGNATURE PREDICTIVE OF INSULIN RESISTANCE

Dr Rima Chaudhuri
UNIVERSITY OF SYDNEY, NSW

BIOGRAPHY

Rima completed her PhD in Bioinformatics and Computational Drug Design in the University of Illinois at Chicago in 2010. Soon after, she moved to the Barcelona Supercomputing Center and Institute for Research in Biomedicine as a postdoctoral candidate in Spain to work on methodology development. Next, she moved to Sydney, Australia and joined the Garvan Institute for systems biology and bioinformatic research in Diabetes and Obesity.

ABSTRACT
Insulin resistance (IR) is perhaps the best predictor of future development of type 2 diabetes (T2D). Identification of molecular signatures that can identify individuals with higher risk of developing T2D could enable early stage intervention. Gene expression data could provide an ideal tool for identification of such a molecular signature, but human gene expression (GE) data is inherently noisy and highly variable. To overcome the lack of power, we integrated human data with more stable data from a model organism and thereby developed a cross-species (human-mouse) analysis platform. We applied this new approach to identify a gene signature that could differentiate insulin resistant from insulin sensitive individuals with improved prediction accuracy (~20% increase) compared to standard clinical measures (OCM). This signature also identified beta-catenin and JAK1 as novel 'connection hubs' to the Insulin Signaling Pathway whose GE pattern highly correlated to extensive metabolic phenotypic measures of insulin sensitivity in humans. Inhibiting these proteins impaired insulin-stimulated glucose uptake in L6 myotubes, confirming their role in insulin action. These data indicate the potential utility of using systems biology approaches to segregate individuals early based on differential diabetes risk.

WEDNESDAY
15 October 2014

SESSION 7

FUNCTIONAL GENOMICS

CHAIRS: KAYLENE SIMPSON AND AMEE GEORGE

0900 – 0945

EXPLORING IMMUNE SIGNALING SYSTEMS WITH HIGH THROUGHPUT, HIGH CONTENT SCREENING

Dr Iain Fraser
NATIONAL INSTITUTES OF HEALTH, USA

BIOGRAPHY
Iain Fraser is Chief of the Signaling Systems Unit in the Laboratory of Systems Biology at the National Institute of Allergy and Infectious Disease, National Institutes of Health. He received his B.S. in biochemistry from Heriot-Watt University, Edinburgh, and his Ph.D. in biochemistry from Imperial College, University of London. He was a Wellcome Trust International postdoctoral fellow at the Vollum Institute and then co-director of the Alliance for Cellular Signaling (AfCS) molecular biology group at the California Institute of Technology. He was appointed as an Investigator at the NIH in 2008. His research has focused on the mechanistic basis of cellular signaling, both in G protein signaling systems and more recently in Toll-like receptor signaling in innate immune cells. He applies systems biology approaches to decipher how mammalian cells integrate stimuli in a complex environment to ensure context-dependent cellular responses. This is vital to understanding how a breakdown in

information processing through cell-surface receptors and their linked signal transduction pathways leads to human disease. Dr.Fraser has developed sophisticated approaches for RNAi-based perturbation analysis of immune cell signaling, as well as cloning platforms and plasmid repositories for high-throughput imaging and cell biological applications.

ABSTRACT

Modern technology now allows the analysis of immune responses and host-pathogen interactions at a global level, across scales ranging from intracellular signaling networks to individual cell behavior to the functioning of a tissue, an organ, and the whole organism. The challenge is not only to collect the large amounts of data such methods permit, but also to organize the information in a manner that enhances our understanding of how the immune system operates or how pathogens affect their hosts. Our laboratory is part of a systems biology program at the NIH whose goal is to develop detailed quantitative models that can be used to predict the behavior of a complex biological system, whose properties help explain the mechanistic basis for physiological and pathological responses to infection or vaccination and can be used to design new therapies or vaccines. We use high-throughput genome-scale technologies (such as siRNA screening and profiling of RNA expression and protein modification) to broadly characterize the cellular response to pathogenic stimuli. I will describe data from ongoing projects in our lab including:

- Genome-wide RNAi screens to characterize signaling network topology in hematopoietic cells and to identify the "parts list" of cellular components involved in immune cell responses.
- Development of software applications for comprehensive analysis of siRNA screening data.
- Use of the above methods to characterize the response of immune cells to defined stimuli, such as single and combined pathogenic ligands, and intact pathogens, such as bacteria and viruses.

0945 – 1015

SYNTHETIC LETHAL TARGETING OF E-CADHERIN-DEFICIENT CANCERS

Professor Parry Guilford

UNIVERSITY OF OTAGO, NEW ZEALAND

BIOGRAPHY

Prof. Parry Guilford is a Principal Investigator in the Cancer Genetics Laboratory, University of Otago, the Director of the Centre for Translational Cancer Research, and the Chief Scientific Officer and co-founder of Pacific Edge Ltd (PEL). He completed his MSc at Otago in 1983, and his PhD at Cambridge University in 1989. His current research interests include the genetics of inherited and sporadic cancers, in particular stomach cancer. Other active research involves the development of genomic-based diagnostic tools for early cancer detection and the application of tumour gene expression and mutation profiling to the optimised selection of chemotherapeutics for cancer patients.

ABSTRACT

Downregulation of the E-cadherin gene (*CDH1*) through either mutation or epigenetic silencing is the hallmark of diffuse gastric cancer and lobular breast cancer. We propose that this loss creates vulnerabilities in the cancer cell that can be exploited with drugs ('synthetic lethal' interactions). To identify the vulnerabilities created by E-cadherin loss, we have conducted a genome-wide siRNA knockdown screen, a 4000 compound known drug screen and a screen of the WEHI's 114,000 compound WECC library in isogenic MCF10a cell lines with and without E-cadherin expression. The functional screen has shown that GPCR signalling proteins are highly enriched amongst the candidate synthetic lethal proteins, as well as many protein classes associated with microtubule and cystoskeletal function. Drug classes that show increased activity against the E-cadherin-deficient cells include the JAK inhibitor LY2784544, the tyrosine kinase inhibitor Crizotinib and several HDAC inhibitors. The WECC library screen has identified a further 88 novel compounds that show synthetic lethality.

This finding paves the way for targeted use of several existing cancer drugs as well as the development of novel drugs and drug combinations for the chemoprevention and treatment of E-cadherin-negative cancers.

1015 – 1030

REPLICATION TIMING IN PROSTATE CANCER

Dr Nicola Armstrong
UNIVERSITY OF SYDNEY, NSW

BIOGRAPHY

After obtaining her PhD in Statistics from UC Berkeley, Nicola moved to the Netherlands, spending time as a post doc before moving to the Netherlands Cancer Institute in Amsterdam as a senior statistician. Since returning to Australia, she spent three years at the Garvan Institute as a senior bioinformatics officer in the Epigenetics Program before moving to the University of Sydney where she is a member of the Statistics research group.

ABSTRACT

DNA replication, during S-phase of the cell cycle, is a tightly regulated developmental process. In a given cell type a specific genomic locus will always replicate at the same time point, a quality known as "replication timing". In order to investigate if an altered replication program has associated changes in the epigenomic landscape, we carried out "Repli-seq" on both normal and cancer prostate cell lines. This technique generates the replication times for all genomic loci. Linking the replication timing data with epigenetic data on histone modifications and methylation levels, we found that epigenetic gene activation and repression at both individual genes and at domains is associated with a shift in replication timing between normal and cancer cells. Additionally, genome-wide hypomethylation, which commonly affects cancer DNA, was found to occur exclusively at DNA that replicates late in both normal and cancer cells.

1030 – 1045

LONG NON-CODING RNA PATHOBIOLOGY IN COLORECTAL CANCER DEVELOPMENT AND PROGRESSION

Dr Sheng Liu
WALTER AND ELIZA HALL INSTITUTE, VIC

BIOGRAPHY

Sheng Liu received his PhD in Molecular Biology from the University of Melbourne in 2013. During his PhD, he studied the regulation of the transforming growth factor-beta (TGF-Î²) signalling pathway with specific focus on the regulation of TGF-Î² receptors. Since early 2013, he has been a postdoc at the Walter and Eliza Hall Institute of Medical Research and is running a high-throughput siRNA-based functional study on human colorectal cancer cell lines to study long non-coding RNA pathobiology.

ABSTRACT

Colorectal cancer (CRC) is the second most common cancer in Australia, affecting 1 in 20 people during their lifetime. In patients presenting with non-metastatic disease the primary cancer can often be resected, but many individuals remain at risk of recurrence. Outcomes for patients who develop metastatic disease are very poor, with a median overall survival of only 24 months, despite an increasing number of chemotherapy and targeted therapy options. LncRNAs are non-protein coding transcripts >200 nucleotides in length that are emerging as important regulators of gene and protein expression in the human genome. To date, several lncRNAs have been implicated in the development of human colorectal cancer (CRC), but a genome-wide survey of lncRNA deregulation with systematic annotation of function is lacking. We hypothesize that discovery of lncRNAs with critical roles in human CRC development and progression can be achieved through integration of genome-wide transcriptome analyses on primary cancers with high-throughput siRNA-based functional studies on human CRC cell lines. In collaboration with the Victorian Centre for Functional Genomics (VCFG), we are the first Australian group to

gain access to the newly developed Human Lincode siRNA Library (Dharmacon), which comprises pools of 4 individual siRNAs against 2,231 validated human lncRNA targets. In a proof-of-principle study, over 250 lncRNAs were detected to be differentially expressed in 11 primary cancers, and many of these were found to impact on CRC cell proliferation consistently by using a panel of 4 CRC cell lines in a siRNA-based knock-down screen.

SESSION 8

SINGLE MOLECULE SEQUENCING

CHAIR: DAVID LOVELL

1115 – 1145

RESOLVING GENOMIC AMBIGUITY: INSIGHTS INTO HUMAN GENOME SEQUENCING AND PATHOGEN SURVEILLANCE USING VERY LONG READS

Dr Robert Sebra

ICAHN SCHOOL OF MEDICINE, MOUNT SINAI INSTITUTE FOR GENOMICS AND MULTISCALE BIOLOGY, NEW YORK, USA

SPONSORED BY

BIOGRAPHY

Robert (Bobby) has extensive experience in using PacBio for human and bacterial genome sequencing. He has also implemented the technology in a clinical setting

ABSTRACT

Whether surveying the human genome for pathologically relevant complex structural variation or investigating various microbes related to infectious disease, long read length sequencing is rapidly enabling more robust references that include previously unachievable domains. Recent headway in single molecule real time (SMRT) sequencing has enabled faster, more comprehensive diagnostics assays. Here, we focus specifically on those methods related to uncovering genetic structural variation towards a more comprehensive sequencing pipeline. Despite the progressive advances in next-generation sequencing technologies over the past decade, a variety of hard to sequence repetitive human genetic loci remain

ambiguous, including many exonic regions with expected biological function. We present data resolving numerous structural variants from whole-genome sequencing data achieved using the RSII platform as well as methods for targeting repetitive genetic regions using long-range, high fidelity PCR in tandem with SMRT sequencing. These results demonstrate both a high N50 reference alongside the potential for diagnostics that better assess genetic variation in pathologically relevant, repeat rich regions and that explore the relationship of structurally complex, non-coding domains using long reads. Further, these results reinforce the pressing need to derive more comprehensive and accurate human genome references to expand their clinical utility beyond that achieved with traditional, shorter read length NGS technologies.

Similarly, genetic diversity among infectious disease isolates spans a multitude of unique pathological structural and epigenetic variants that identify factors related to virulence, origin, and resistance. Addressing the need to finish microbial genomes to better identify these factors, we present a rapid method to isolate genomic DNA and generate read lengths with N50 values as high as 12,000 base pairs using SMRT sequencing. Through a series of studies, a demonstration of how readlengths as long as 60,000 base pairs have enabled the ability to assemble complex structural domains, identify structural insertion elements, phase accessory genomes such as plasmids, and how these reads are used to better understand microbial genetics and transmission will be discussed. One specific example from a solid organ transplantation case will detail the comparison of Staphylococcus aureus isolates in donor and recipient and how microbial finishing can be used to further improve our understanding of the epidemiology of bacterial transmission and the risk of adverse patient outcomes when a compromised organ is considered. Additionally, better characterization of community isolates may improve surveillance, reducing the number of hospital-acquired cases and minimizing the risk of outbreak. Various microbial case studies will be presented to illustrate this development.

SESSION 9

METAGENOMICS

CHAIRS: TORSTEN SEEMANN AND MARK SCHULTZ

1345 – 1415

TOWARD ROUTINE CLINICAL SURVEILLANCE OF THE HUMAN MICROBIOTA

Associate Professor Aaron Darling
UNIVERSITY OF TECHNOLOGY SYDNEY, NSW

BIOGRAPHY

Aaron Darling is an Associate Professor in Computational Genomics and Bioinformatics in the UTS Faculty of Science's ithree institute. He has over a decade of experience developing computational methods for comparative genomics and evolutionary modeling and in 2013 moved from the University of California-Davis to start a computational genomics group at UTS. Darling has led the development of several widely used software packages, including the Mauve software for genome alignment & comparison, the mpiBLAST parallel BLAST software, and the BEAGLE library for GPU accelerated statistical phylogenetics used in BEAST and MrBayes.

He is an active advocate for open access, open source, and open peer review in scientific communication.

ABSTRACT

Bacterial genomes have proven to be remarkably dynamic entities, containing systems that incorporate foreign gene content and an ability to maintain substantial population-level genetic diversity. Isolates of the same species can differ markedly in their gene content, with the collection of genes common to all members of the

species representing 60% or less of a strain's genome. The extent of variation both in gene content and point mutations is dependent on population genetic processes including population demographics and recombination.

Meanwhile, the human body plays host to a complex microbial ecosystem, the development of which likely begins before birth. After birth, the human microbiota continues to develop, and emerging evidence suggests that the trajectory of this development is associated with long term health outcomes. Routine monitoring of the development of microbial ecosystems in newborns (or other environments) using metagenomic methods is currently extremely challenging and expensive. I will describe some recent technological advances that could enable routine sequencing and computational analysis of hundreds of metagenomes, and demonstrate their application to samples of an infant fecal microbiome during the first months of life. In this study forty-five samples were subjected to metagenomic sequencing. The microbial community structure in the resulting data were analysed with a new approach called phylogenetic Edge Principal Component Analysis (Edge PCA) that can identify which lineages in a phylogeny explain the variation among the samples. Major life events in the infant appear to be associated with dramatic change in the gut microbiota. Strain or substrain-level genomic variation is common in many ecosystems and also appears in the infant gut. This strain-level variation remains especially difficult to characterise. I will discuss the potential for a new technique called metagenomic Hi-C to help characterise the strain-level variation in microbiome samples.

1415 – 1445

THE INFANT AIRWAY MICROBIOME

Dr Kathryn Holt

BIO21 INSTITUTE, UNIVERSITY OF MELBOURNE, VIC

BIOGRAPHY

Kat completed her PhD in 2009 at the Wellcome Trust Sanger Institute and University of Cambridge, studying bacterial comparative genomics and genomic epidemiology. In 2010 she returned to Australia to take up an NHMRC ECR Fellowship in the Department of Microbiology and Immunology at the University of Melbourne. While working as a research fellow, she undertook a Masters in Epidemiology at the university, graduating in 2011. In 2012 she was awarded a Rod Rickards Fellowship from the Australian Academy of Science to spend time at the Pasteur Institute in Paris, and obtained several NHMRC Project Grants to study various aspects of microbial genomics in relation to human health. In December 2012 she became a lab head in the Department of Biochemistry and Molecular Biology at the University of Melbourne. In 2013 she was awarded a L'Oréal Australia Women in Science Fellowship and is now an NHMRC Career Development Fellow.

ABSTRACT

Respiratory infections are common in young infants and are a major cause of morbidity and mortality. In the past decade, respiratory infections during infancy have been recognized as an important factor driving the development of asthma during childhood. In addition, bacterial colonisation of the airways has been shown to influence viral infection and asthma development. We set out to characterize the bacterial composition of the nasopharyngeal microbiome (NPM) in a cohort of 234 infants at high risk of allergy, which was previously established to investigate the role of respiratory infection in asthma development. The infant NPMs had a simple structure with six major types and were subject to dynamic changes during the first year of life, influenced by childcare, siblings, breastfeeding, season, antibiotics and incident respiratory infections.

Early colonization with Streptococcus was associated with the development of allergic asthma by age five. Streptococcus, Haemophilus and Moraxella were associated with the presence and severity of symptoms of respiratory infection, regardless of the presence of virus. Moraxella was also a remarkably stable colonizer of the infant airway, whose presence was associated with increased incidence and severity of infections with respiratory syncitial virus (RSV).

1445 – 1500

MAPPING SHARED MARKERS ACROSS A POLYNESIAN POPULATION IN THE GENOMIC AGE

Ms Sophia Cameron-Christie
UNIVERSITY OF OTAGO, NEW ZEALAND

BIOGRAPHY

Sophia Cameron-Christie is a PhD student in the Clinical Genetics Lab at the University of Otago, New Zealand, following an honours degree in genetics at Otago. She has worked to identity mutations for rare, Mendelian diseases with next generation sequencing and to use both array and NGS data for linkage analysis and haplotype discovery. She is interested in studying population history through genomic data.

ABSTRACT

Identifying susceptibility variants for multifactorial human traits can be complicated depending on the understanding of the genetic background on which it arises. This is especially difficult in small, under-studied populations without readily-available control cohorts. In such populations, low-penetrance variants in complex disease can be challenging to confidently identify when normal variation and allele frequencies are uncharacterised. We present an investigation of biliary atresia (BA), a usually sporadic malformation of the biliary tree. Worldwide, BA leads to half of all paediatric liver transplants, and is always fatal without major surgical intervention. The causes of BA remain unknown; autoimmune processes and genetic background may both play a part. In Maori and Polynesian populations the incidence is elevated three-fold compared to Europeans. We have identified a large Maori family (iwi) exhibiting an extremely elevated incidence of BA (1:100–300). To circumvent some of the problems in studying complex traits in Maori we have adopted a non-parametric, family-based approach to localise a presumptive genetic factor conferring this susceptibility. Assuming a single, segregating susceptibility factor contributes to BA in our family cohort, we have used the software Beagle, Germline and in-house methods to examine haplotype sharing between affected pairs across this family. We have constructed maps of long, Identical-By-State (IBS) segments across the collective genomes of affected individuals without reliance on pre-existing assumptions about allele frequencies, a definitive inheritance model or exclusion of unidentified phenocopies. Further work aims to integrate rare variation from next generation sequencing data to further resolve shared haplotypes and define candidate regions in which a susceptibility variant could lie.

1500 – 1515

PORE PERFORMANCE? BENCHMARKING THE OXFORD NANOPORE TECHNOLOGIES MINION FOR METAGENOMICS AND CONTIG ASSEMBLY.

Dr Ken McGrath

THE AUSTRALIAN GENOME RESEARCH FACILITY, QLD

BIOGRAPHY

Ken McGrath is the manager of the Brisbane Lab of the Australian Genome Research Facility. He completed his undergraduate degree with honours in 2001 at QUT working with the plant biotechnology group on developing transgenic bioreactors, and transitioned to UQ for his PhD work investigating the genetic regulation of plant defence responses to disease, in collaboration with CSIRO and the CRC for Tropical Plant Protection. Following this, his post-doctoral research with the Schmidt and Schenk labs at UQ involved examining the transcriptomes of mixed microbial communities in industrial and agricultural settings. In 2009, Ken joined the AGRF as sequencing supervisor, and currently helps manage submissions and workflows on a range of next-generation sequencing platforms.

ABSTRACT

The Oxford Nanopore Technologies (ONT) MinION is a revolutionary new device capable of directly sequencing single DNA molecules via translocation through a protein pore. The Australian Genome Research Facility (AGRF) was chosen to take part in the MinION "early access" sequencing program, which allowed us to receive and trial the platform prior to general release.

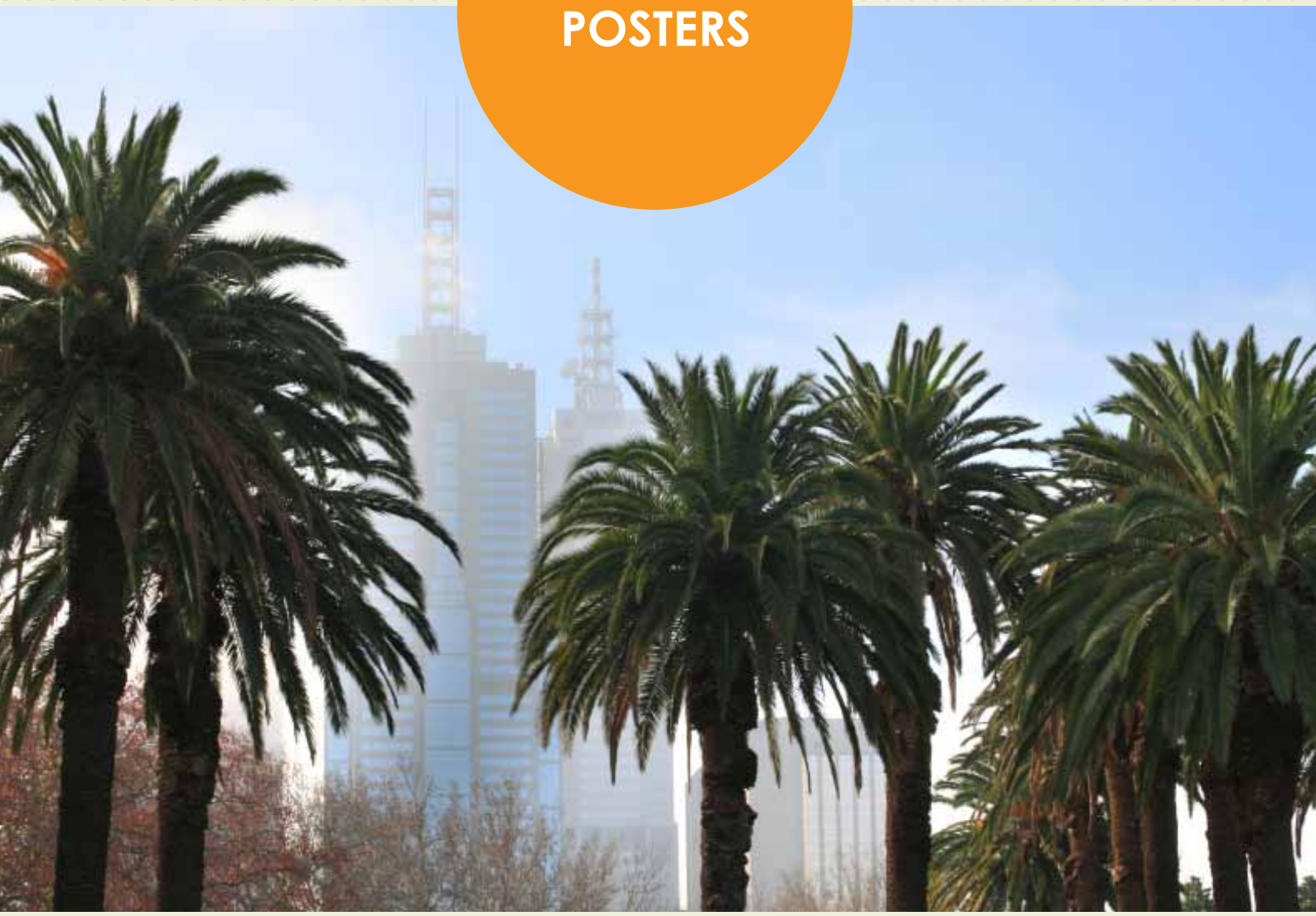
We have performed several sequencing runs on this device, including whole-genome shotgun sequencing, as well as targeted long-amplicon sequencing.

These initial runs generated between 1000-5000 reads events per 6hr run, with read lengths ranging from 5bp to over 42,000bp. Analysis of these reads showed a high error rate when compared to the target sequence with a particular tendency to form gaps (both as insertions and deletions).Correct and incorrect calls tended to form clusters along the reads, suggesting the reads can enter an "error-prone state" during certain sections or phases of the read. Both the systematic and random nature of these errors was also assessed.

We assessed the utility of the MinION platform in 2 applications where long sequencing reads are particularly useful – microbial community profiling and de novo assembly. We trialled the platform to identify bacteria from a known artificial microbial community based on sequencing the 16S region, and compared these results with alternate methods available on established sequencing platforms such as the Roche 454 and Illumina MiSeq. We also assessed the utility of MinION reads for scaffolding contigs in the de novo assembly of short-read sequencing data sets.

This work highlights the strengths and weaknesses of this emerging technology, and provides an insight into what the future of single-molecule sequencing holds.

POSTERS



POSTER PRESENTATIONS

MONDAY 13 OCTOBER 2014

#	POSTER TITLE	PRESENTING AUTHOR	
1	TARGETED MASSIVELY PARALLEL SEQUENCING OF THE MITOCHONDRIAL PROTEOME	Hayley Mountford	O
2	MODELLING CELLULAR CONTEXT TO MAP KINASE-SUBSTRATE PHOSPHORYLATION EVENTS	Ralph Patrick	O
3	STEMFORMATICS: A USER-FRIENDLY DATABASE OF WELL-CURATED BIOLOGICAL DATA	Tyrone Chen	
4	STATISTICAL ANALYSIS FOR ALLEIE- AND HOMEOLOG-SPECIFIC EXPRESSION WITH RNA-SEQ	Satoru Akama	O
5	INTRODUCING ENRICH AND THE INTERFEROME BIOINFORMATICS TOOLS, USING PROMOTER ANALYSIS TO UNDERSTAND CYTOKINE SIGNALLING PATHWAYS	Helen Cumming	
6	ANALYSING TANDEM REPEATS BY HIGH THROUGHPUT SEQUENCING	Devika Ganesamoorthy	O
7	BIOINFORMATICS APPROACHES, CHALLENGES AND PERFORMANCE OF DIVERSITY PROFILING ON THREE NEXT GENERATION SEQUENCING PLATFORMS	Naga Kasinadhuni	
8	THE EXTENSION OF VOOM TO TEST PREFERENTIALLY SPLICED EVENTS IN RNA-SEQ COUNT DATA	Charity Law	
9	VARIANTGRID: A GRAPHICAL INTERFACE FOR A SCALABLE VARIANT DATABASE	David Lawrence	O
10	A COMPARISON STUDY OF CPU AND GPU-BASED SHORT READ ALIGNMENT TOOLS	Xi Li	O
11	BRAEMBL GALAXY CONNECT SUITE: HARNESSING THE DATA AND COMPUTE RESOURCES OF THE POWERFUL BRAEMBL COMPUTER CLUSTER THROUGH THE FRIENDLY GALAXY INTERFACE	Webber W.P. Liao	O
12	COMBINE, A GROUP AIMED AT STUDENTS AND EARLY-CAREER RESEARCHERS IN BIOINFORMATICS, COMPUTATIONAL BIOLOGY AND SYSTEMS BIOLOGY	Andrew Lonsdale	O
13	IN SILICO GENE PRIORITISATION WITH BRAINGEP	Vesna Lukic	
14	CONTROLLING THE ERROR RATE CORRECTLY WHEN INTERSECTING GENE LISTS	Aaron Lun	O

O STUDENT O EARLY CAREER RESEARCHER

#	POSTER TITLE	PRESENTING AUTHOR	
15	INTEGRATING NEXT GENERATION SEQUENCING DATA WITH ADDITIONAL 'OMICS DATASETS	Paul Maclean	
16	RISK-CONSCIOUS CORRECTION OF BATCH EFFECTS: MAXIMISING INFORMATION EXTRACTION FROM HIGH-THROUGHPUT GENOMIC DATASETS	Yalchin Oytam	
17	ONCOCIS : AN ANNOTATION TOOL FOR CIS-REGULATORY MUTATIONS IN CANCER	Dilmi Perera	O
18	DIFFVAR: A NEW METHOD TO TEST FOR DIFFERENTIAL VARIABILITY APPLIED TO DNA METHYLATION IN CANCER AND AGING	Belinda Phipson	O
19	ALLELE FREQUENCY CALIBRATION IN POOLED DNA SAMPLES: A MACHINE LEARNING APPROACH	Ashfaqur Rahman	
20	INTEGRATIVE OMICS ANALYSIS ELUCIDATES ROLE FOR INFLAMMATION AND NEUTROPHIL DEGRANULATION IN GLYCOPROTEIN-ASSOCIATED ALL-CAUSE MORTALITY RISK	Scott Ritchie	O
21	WHY WEIGHT? VARIANCE MODELLING FOR DESIGNED RNA-SEQ EXPERIMENTS	Matthew Ritchie	
22	STATISTICAL METHODS AND SOFTWARE FOR FUNCTIONALLY CHARACTERIZING EVERY SINGLE MUTATION IN YOUR FAVORITE GENE	Alan Rubin	O
23	A STATISTICAL METHOD TO DETECT PATHOGENIC SHORT-TANDEM REPEAT EXPANSIONS IN MASSIVELY-PARALLEL SEQUENCING	Rick Tankard	O
24	OPTIMISING MULTI-CALLER APPROACHES TO SOMATIC VARIANT CALLING	Paul Wang	
25	DIFFERENTIAL EXON USAGE DETECTION WITH RNA-SEQ USING ALTERED COUNT SUMMARISATION OPTIONS	Katrina Bell	
26	WHOLE-GENOME BISULFITE SEQUENCING (WGBS) AN IMPROVED "POST-BISULFITE" CONVERSION LIBRARY CONSTRUCTION METHOD FROM LOW GDNA INPUTS	Leah Alcock	
27	EPIGENETIC CHANGES IN SURVIVORS OF PRETERM BIRTH: EFFECT OF GESTATIONAL AGE AND EVIDENCE FOR A LONG TERM LEGACY	Jeffrey Craig	
28	A COMPARISON OF CONTROL SAMPLES FOR CHIP-SEQ OF HISTONE MODIFICATIONS	Christoffer Flensburg	O
29	EZH2 IS ESSENTIAL FOR LUNG DEVELOPMENT, MAINTENANCE OF EPITHELIAL CELL IDENTITY AND CELL FATE DETERMINATION	Aliaksei Holik	
30	GENOMEWIDE ANALYSIS OF PROTEINS THAT BIND TO DNA AND REGULATE GENE EXPRESSION	Siobhan Hughes	O

O STUDENT O EARLY CAREER RESEARCHER

#	POSTER TITLE	PRESENTING AUTHOR	
31	EPIGENETICS, T-CELLS AND AUTOIMMUNITY: A BIOINFORMATICS STORY	Jovana Maksimovic	ECR
32	MUTATIONS IN LEPR GENE ARE ASSOCIATED WITH REPRODUCTIVE TRAITS IN DAVISDALE SHEEP	Michelle French	
33	EXPLORING THE PHEROMONE GLANDS OF NATIVE NEW ZEALAND LEAFROLLER MOTHS USING DE NOVO TRANSCRIPTS PRELIMINARY INSIGHTS	Amali Thrimawithana	ECR
34	A BOVINE GENE EXPRESSION ATLAS	Christy Vander Jagt	
35	COMPARATIVE TRANSCRIPTOMICS OF ACACIA ACUMINATA PARASITISED VS. NON-PARASITISED BY THE HEMIPARASITIC SHRUB SANTALUM SPICATUM	Bjorn Dueholm	STUDENT
36	DYNAMIC ACTIVITIES OF NOVEL TRANSCRIPTS IN LEAF AND SHOOT APICAL MERISTEM OF SOYBEAN DURING FLORAL TRANSITION	Chol-Hee Jung	ECR
37	TOWARDS GENOMIC SELECTION FOR ESSENTIAL OIL YIELD IN EUCALYPTUS AND MELALEUCA	David Kainer	STUDENT
38	EFFICIENT GENOTYPING-BY-SEQUENCING ENABLES CONSTRUCTION OF THE FIRST LINKAGE MAP FOR THE NON-MODEL LEGUME FORAGE SPECIES, BITUMINARIA BITUMINOSA VAR. ALBOMARGINATA (TEDERA)	Jafar S. Jabbari	
39	MICROSATELLITE MARKER-BASED IDENTIFICATION AND GENETIC RELATIONSHIPS OF MILLENNIUM OLIVE CULTIVARS IN TUNISIA	Mnasri Sameh	ECR
40	ASSEMBLY AND ANALYSIS OF A MALE STERILE RUBBER TREE MITOCHONDRIAL GENOME REVEALS DNA REARRANGEMENT EVENTS AND A NOVEL TRANSCRIPT.	Jeremy Shearman	ECR
41	SINGLE CELL-TYPE ANALYSIS REVEALS UNIQUE PATTERNS OF DNA METHYLATION IN THE ROOT MERISTEM	Tim Stuart	
42	HIGH THROUGH-PUT SEQUENCING GUIDES THE DESIGN OF DNA MINOR GROOVE BINDING SMALL MOLECULE WITH IMPROVED BINDING SPECIFICITY	Anandhakumar Chandran	
43	END-TO-END SAMPLE QUALITY CONTROL FOR NEXT GENERATION SEQUENCING LIBRARY PREPARATION AND SURESELECT TARGET ENRICHMENT ON THE AGILENT 2200 TAPESTATION SYSTEM	Arunkumar Padmanaban	
44	EVALUATION OF FLUIDIGM SNPTRACE™ PANEL, WITH A COMPARISON TO DNA QUAL AND BIOANALYZER FRAGMENT ANALYSIS, FOR DNA QUALITY ASSURANCE WITHIN THE TUMOUR BANK PROGRAM	Li Zhou	ECR
45	A BETTER WAY - AMPLISEQ EXOME FOR CLINICAL RESEARCH	Richard Allcock	

POSTER PRESENTATIONS

TUESDAY 14 OCTOBER 2014

#	POSTER TITLE	PRESENTING AUTHOR	
46	EXOMIC MOLECULAR SIGNATURES SEGREGATE ORAL EPITHELIAL LESIONS BASED ON DISEASE SEVERITY	Nigel Bennett	○
47	GENOMIC EVOLUTION OF HEPATOCYTES TOWARDS HEPATOCELLULAR CARCINOMA	Magdalena Budzinska	
48	DETECTING TUMOUR DNA IN BLOOD PLASMA USING ALLELIC IMBALANCE FROM READ COUNTDATA	Wenhan Chen	○
49	THE ROLE OF MEK SIGNALING IN DRIVING MESENCHYMAL TRANSITION IN CELL LINE MODELS OF HUMAN BREAST CANCER METASTASIS	Erik Thompson	
50	DECIPHERING GENOMIC EVOLUTION IN HETEROGENEOUS TUMOURS	Vincent Corbin	○
51	MIRNA-TARGET GENE RELATIONSHIP ANALYSES IDENTIFY KEY ONCOGENES FOR KIDNEY RENAL||CLEAR CELL CARCINOMA PROGRESSION	James Doecke	
52	AMPLICON-BASED TARGETED RESEQUENCING IDENTIFIES NOVEL GENETIC RISK FACTORS FOR SARCOMA AND AN ASSOCIATION BETWEEN GENETIC LOAD AND AGE OF ONSET IN SARCOMA PATIENTS	David Goode	○
53	MICRORNA PROGNOSTIC SIGNATURES IN METASTATIC MELANOMA	Kaushala Jayawardana	○
54	NEXT GENERATION SEQUENCING OF HAEMATOLOGICAL DISEASES IN A CLINICAL SETTING A PILOT STUDY	Anabel Kearney	
55	MOLECULAR PROFILE OF ORAL MUCOSAL LESIONS CHARACTERISED BY DIRECT TISSUE AUTOFLUORESCENCE	Farzaneh Kordbacheh	○
56	EXOME SEQUENCING REVEALS BCL6 CO-REPRESSOR (BCOR) AS A FREQUENTLY MUTATED TUMOR SUPPRESSOR GENE IN E□-MYC LYMPHOMA	Marcus Lefebure	○
57	GENOME-WIDE MULTIMARKER MODEL TO PREDICT THE RISK OF BLADDER CANCER RECURRENCE	Evangelina Lopez De Maturana	
58	STRATIFICATION OF THE CANCER GENOMIC LANDSCAPE THROUGH THE DEVELOPMENT OF A PAN-CANCER GENETIC SCREENING PANEL	Mark McCabe	
59	STRUCTURAL VARIATION IN PANCREATIC CANCER GENOMES	Kelly Quek	○

○ STUDENT ○ EARLY CAREER RESEARCHER

#	POSTER TITLE	PRESENTING AUTHOR	
60	COLORECTAL CANCER METHYLATION BIOMARKERS	Jason Ross	
61	SELECTION RATHER THAN MUTATION DRIVES INCREASED PLATINUM RESISTANCE IN A PATIENT DERIVED XENOGRAFT MODEL OF OVARIAN CANCER	Matthew Wakefield	
62	WHOLE EXOME SEQUENCING IDENTIFIES A RECURRENT RQCD1 P131L MUTATION IN CUTANEOUS MELANOMA	Stephen Wong	○
63	IMPLEMENTING CLINICAL GENOMICS FOR DIAGNOSTICS AND RESEARCH	Mark Cowley	
64	GENOME CHARACTERISATION OF THE CELL LINE U937 PROVIDES FURTHER EVIDENCE FOR CENTROMERE CAPTURE AND DEMONSTRATES THE USE OF MULTIPLE GENOME TECHNOLOGIES TO EXAMINE EVOLUTION IN CELL LINES	Ruth MacKinnon	
65	LIBRARY PREPARATION, TARGETED-HYBRIDISATION CAPTURE AND SEQUENCING OF LIMITING AMOUNTS OF FFPE SAMPLES	Gisela Mir	
66	A CLINICAL READY SYSTEM FOR SCALABLE GENOMICS ANALYSIS	Mahtab Mirmomeni	
67	NEXT-GENERATION SEQUENCING - BEYOND VARIANTS	John Pearson	
68	CHALLENGES TO CNV DETECTION IN CLINICAL SETTINGS USING TARGETED HIGH THROUGHPUT SEQUENCING DATA	Simon Sadedin	○
69	A HANDHELD QPCR DEVICE FOR POINT-OF-CARE AND IN-FIELD DIAGNOSTICS	Jo-Ann Stanton	
70	GENERATING ACTIONABLE KNOWLEDGE FROM COMPLEX GENOMIC DATA FOR PERSONALISED CLINICAL DECISIONS.	Siamak Tafavogh	○
71	NEXT GENERATION SEQUENCING FOR PHARMACOGENETIC SCREENING	Kim Ton	○
72	CAUSES OF MULTIPLE SCLEROSIS: A FUNCTIONAL GENOMICS APPROACH	Margaret Jordan	○
73	HOW DOES CHROMOSOME 21 TRISOMY ALTER THE GENE REGULATORY NETWORK OF NEURAL DIFFERENTIATION?	Nicholas Matigian	
74	EPILEPSY: NEW GENES AND OLD PATHWAYS IN A COMPLEX DISORDER	Leanne Dibbens	
75	CROSS-SPECIES SYSTEMS ANALYSIS OF A METABOLIC DISEASE TO FIND A GENE SIGNATURE PREDICTIVE OF INSULIN RESISTANCE	Rima Chaudhuri	

#	POSTER TITLE	PRESENTING AUTHOR	
76	IMPERFECT CENTERED MIRNA BINDING SITES ARE COMMON AND FUNCTIONAL	Nicole Cloonan	
77	DERIVING MYCOBACTERIUM TUBERCULOSIS SPOLIGOTYPES FROM NGS DATA	Thomas Conway	
78	GENOTYPING MICROSATELLITES IN NEXT-GENERATION SEQUENCING DATA	Harriet Dashnow	O
79	IDENTIFYING DISEASE-ASSOCIATED PATHWAYS IN GENOME-WIDE ASSOCIATION STUDIES VIA A NETWORK-BASED KERNEL APPROACH	Saskia Freytag	
80	GENE DISCOVERY IN LARGE MULTIPLEX FAMILIES WITH AUTISM SPECTRUM DISORDERS	Miriam Fanjul Fernandez	O
81	THE GENETIC BASIS OF LEIGH SYNDROME, THE MOST COMMON MITOCHONDRIAL DISORDER AFFECTING CHILDREN: A COHORT STUDY	Nicole Lake	
82	COMPUTATIONAL DECONVOLUTION OF GENOME WIDE EXPRESSION DATA FROM PARKINSON'S AND HUNTINGTON'S DISEASE TISSUES USING POPULATION-SPECIFIC EXPRESSION ANALYSIS.	Victoria Perreau	
83	MUTATIONS IN THE MTORC1 REGULATOR DEPDC5 ARE A MAJOR CAUSE OF LESIONAL AND NON-LESIONAL FOCAL EPILEPSY	Michael Ricos	
84	INFERRING IDENTITY BY DESCENT ON THE X CHROMOSOME	Lyndal Henden	O
85	GENOTYPING-BY-SEQUENCING (GBS): COMPARISON TO SNP CHIPS AND WHOLE GENOME SEQUENCING	Tracey Van Stijn	
86	EXTENSIVE GENETIC CO-REGULATION DRIVES HIGHLY CORRELATED MODULAR STRUCTURES IN TRANSCRIPTOMIC DATA	Anita Goldinger	O
87	THE DISCOVERY OF NOVEL LINCRNA WITH POTENTIAL ROLES IN MESC DEVELOPMENT	Beth Signal	
88	DIVERSIFICATION OF INNATE IMMUNE RESPONSES BY TRANSCRIPTIONAL MECHANISMS	Suzanne Butcher	O
89	COMPLEX CODING AND NONCODING TRANSCRIPTIONAL DYNAMICS IN EARLY STEM CELL DEVELOPMENT	Brian Gloss	
90	DETECTING DIFFERENTIAL EXON USAGE IN RNA-SEQ	Yunshun Chen	O

#	POSTER TITLE	PRESENTING AUTHOR	
91	CHARACTERIZING CHANGES IN EPIGENOME AND TRANSCRIPTOME DURING MAMMARY STEM CELL DIFFERENTIATION	Wei Shi	
92	RNA-SEQ ANALYSIS OF LIP TISSUE FROM A MOUSE MODEL OF WILLIAMS-BEUREN SYNDROME	Susan Corley	○
93	DISCOVERY OF ALLELE SPECIFIC EXPRESSION IN RNASEQ DATA FROM EIGHTEEN BOVINE TISSUES	Amanda Chamberlain	
94	STRANDED VERSUS NON-STRANDED RNA-SEQ	Sonika Tyagi	

Monday

13 OCTOBER

POSTER 1

Hayley Mountford , Alison Compton , Elena Tucker, Sarah Calvo , Bas Wanschers , Radek Szklarczyk , Michael Ryan , Martijn Huynen , Vamsi Mootha , David Thorburn

TARGETED MASSIVELY PARALLEL SEQUENCING OF THE MITOCHONDRIAL PROTEOME

ABSTRACT

Inherited defects in mitochondrial oxidative phosphorylation (OXPHOS) are the most common inborn error of metabolism, affecting at least 1 in 5000 live births, predominantly affecting organs with high-energy consumption. OXPHOS diseases are notoriously difficult to diagnose, as they show extreme clinical heterogeneity, presenting at any age and with any level of severity, and typically impact on multiple organ systems.

They are also genetically heterogeneous with over 140 mitochondrial DNA and nuclear DNA encoding genes associated with OXPHOS disease. Despite this, ~50% patients with OXPHOS disease do not have a molecular diagnosis.

We developed a targeted DNA capture and massively parallel sequencing method to detect variants within 1034 genes encoding the mitochondrial proteome (MitoExome). We applied this MitoExome sequencing approach to a group of 44 unrelated patients with clinical and biochemical evidence of severe OXPHOS disease. Molecular diagnoses were provided for 13 patients with mutations in 10 known OXPHOS disease genes. This approach identified 7 novel confirmed disease genes in 9 patients; *NDUFB3*, *AGK*, *MTFMT*, *EARS2*, *LYRM4*, *COA6* and *UQCC2*.

We identified *UQCC2* as a novel disease gene in a consanguineous Lebanese patient with severe CIII deficiency, and demonstrated aberrant splicing. We showed that UQCC2 co-purifies with binding partner UQCC1 and have proven that both are required for early CIII assembly (Tucker et al, PLoS Genet 2013).

MitoExome sequencing shows promise as a research tool for discovering novel disease loci and as a noninvasive diagnostic approach for patients with clinical evidence of mitochondrial disease, however proof of pathogenicity still remains challenging.

POSTER 2

Ralph Patrick, Kim-Anh Le Cao, Bostjan Kobe, Mikael Boden

MODELLING CELLULAR CONTEXT TO MAP KINASE-SUBSTRATE PHOSPHORYLATION EVENTS

ABSTRACT

The determinants of kinase-substrate binding can be found both in the substrate sequence, and the surrounding cellular context. Cell cycle progression, interactions with mediating proteins and even prior phosphorylation events are necessary for kinases to maintain substrate specificity. While much work has focussed on the use of sequence based methods to predict kinase-specific phosphorylation sites, lack of specificity in many kinase binding motifs means that sequence methods for predicting kinase binding sites are susceptible to high false-positive rates. Few studies have attempted to leverage the cellular context that kinases operate in; and while context information is readily available in various databases, incomplete coverage and variable certainty means that the integration of context features into a model is non-trivial.

We present a Bayesian network model that can accommodate missing values, seamless combination of heterogeneous data, and to provide flexible options for querying potential kinase substrates. The model integrates known kinase-substrate relationships, protein-protein interactions, and protein abundance across the cell cycle to predict kinase substrates across a variety of human kinase families. Our model shows high prediction accuracy, with a mean AUC of 0.86 across

the kinases tested. When using the model to complement sequence based kinase-specific phosphorylation site prediction, we find that the additional information can greatly increase prediction performance at low false positive levels. Our results demonstrate that the model can account for the short-falls in sequence information and provide a robust description of the regulation of kinase-protein phosphorylation.

POSTER 3

Mr Tyrone Chen, Mr Rowland Mosbergen, Mr Othmar Korn, Associate Professor Christine Wells

STEMFORMATICS: A USER-FRIENDLY DATABASE OF WELL-CURATED BIOLOGICAL DATA

ABSTRACT

The exponential growth of 'big-data' to address biological networks has greatly increased the type and scale of raw data available to the research community. This led to the creation of publicly-accessible computational repositories for biological data, allowing for external re-analyses and further studies on datasets. However, existing databases are driven by data-generators, not data-users. This means that the quality of the experimental paradigm and data is uncurated, and needs to be examined dataset-by-dataset. Data formats vary considerably between platforms, which coupled with their volume makes it difficult to identify datasets relevant to any specific biological question.

To address this, we have developed a biology-centric approach. Stemformatics was designed with the aim of integrating biology, computational methods and statistics in a format intuitively accessible by stem cell researchers. The Stemformatics project currently houses 236 datasets, of which 122 are publicly available. Each dataset has been carefully curated for cell phenotype and experimental design, with rigorous quality control steps. Available datasets comprise gene expression data from a range of platforms including microarrays, next-generation sequencing, proteomics and ChIP-seq platforms. Currently the database contains mouse and human expression data, with support for additional model organisms being planned. The user interface is designed for rapid assessment of the expression of individual or multiple genes across and/or within datasets. Users can store gene lists and share views with collaborators. Stemformatics is funded as part of the ARC special initiative "Stem Cells Australia", has an international community of users and is freely accessible to all academic groups.

POSTER 4

Satoru Akama, Rie Shimizu-Inatsugi, Kentaro Shimizu, Jun Sese

STATISTICAL ANALYSIS FOR ALLEIE- AND HOMEOLOG-SPECIFIC EXPRESSION WITH RNA-SEQ

ABSTRACT

Next-generation sequencers allow us to distinguish expression levels of gene pair in homologous chromosomes, such as paternal- and maternal-specific expressions in diploid species or homeolog-specific in allopolyploid. Recent researches show that the expression balances are greatly changed according to cellular status, causing differentiations and adaptations in various species. However, the lack of its statistical assessment method has prohibited the connection between the expression difference of the gene pair and the understanding of biological functions. For example, while existing research uses Fisher's exact test or chi-square test to check the changes of expression balance, our computational experiments confirmed the high false positive ratio of the detection with traditional statistics because of their no consideration of replicate experiments. To solve the problem, we introduce a new statistical analysis procedure for allele- or homeolog-specific expressions including two new methods. One is to detect the genes whose expression ratios are changed between different environments named HomeoRoq

(http://seselab.org/homeoroq/) [Akama et al., NAR, 42(6):e40, 2014]. The other is the statistically sound detection method of allele- or homeolog-specific silencing based on the estimation of the distribution of genome-wide expression ratio using a mixture model. We applied the procedure to allotetraploid *Arabidopsis kamchatica* derived from diploid *A. lyrata* and *A. halleri* to examine the environmental adaptation in evolutionary process. The results showed that the procedure could reveal that homeolog sensitive genes were highly related to the stress response genes, contributing to the environmental adaptation.

POSTER 5

Dr Helen Cumming, Dr Sam Forster, Dr Alex Finkel, Dr Ross Chapman, Dr Jamie Geering, Dr Kevin Luu, Mr David Hughes, Prof. Paul Hertzog

INTRODUCING ENRICH AND THE INTERFEROME BIOINFORMATICS TOOLS, USING PROMOTER ANALYSIS TO UNDERSTAND CYTOKINE SIGNALLING PATHWAYS

ABSTRACT

Introduction

The interferon cytokine family plays a pivotal role in protecting the body from pathogenic infections and tumour development. The interferon signalling pathway is a tightly regulated and highly structured temporal response, with stimulation activating an organised sequence of transcriptomic events. Here we present two bioinformatics tools that can be used to analyse the transcriptional regulation of the Interferon response: 1) the Interferome v2.0 database (http://interferome.its.monash.edu.au/interferome/home.jspx), which identifies genes regulated by Interferons, and 2) the Enrich custom software, which is a powerful and intuitive tool for identifying transcription factors involved in regulating co-expressed gene sets and thus elucidating signaling pathways.

Methods

The *Interferome* database was queried for genes that were up-regulated in murine cells over discrete time periods following stimulation by type I Interferons. Transcription factors that were significantly enriched ($p<0.05$) among each gene set were identified using the *Enrich* custom bioinformatics tool.

Results

Interrogation of the *Interferome* database identified hundreds of genes that are up-regulated following stimulation by Interferon. Analysis of the associated promoters using *Enrich*, identified clear patterns in the transcription factors binding sites associated with temporally co-expressed genes. Many transcription factors identified are already known to be involved in the Interferon pathway, including STATs and IRFs. The analysis also revealed the additional transcription factors which might provide new insights into Interferon signalling.

Conclusions

By integrating the *Interferome* and *Enrich* bioinformatics resources, it has been possible to explore the transcription factors that regulate different phases of the interferon response and identify new regulatory elements for experimental validation.

POSTER 6

Dr Devika Ganesamoorthy, Dr Minh Duc Cao, Mr Wenhan Chen, Dr Lachlan Coin

ANALYSING TANDEM REPEATS BY HIGH THROUGHPUT SEQUENCING

ABSTRACT

The contribution of genetic variation to complex disease susceptibility has been extensively studied in the recent years. Although the influence of copy number variations and SNPs on complex diseases have been explored in detail, the effect of tandem repeats (TRs) are poorly understood. There are almost 1 million TRs in the human genome, however only few of these regions have been investigated in terms of disease association. This is mainly due to the hyper-variable and multi-allelic nature of TRs which makes it difficult to be tagged by SNPs for genome wide association studies. The recent advances in high throughput sequencing (HTS) provides an opportunity to unlock this type of variation. We have developed a target enrichment based sequencing method to analyse 140 TRs associated with obesity. The targeted TRs range from 2bp to 1800bp in repeat unit length and varies from 2 to 2300 number of copies. A pilot study was performed to assess the feasibility of the targeted sequencing approach. More than 90% of sequence reads mapped to the targeted regions, which includes the TRs and the flanking regions. Greater than 100X coverage was achieved in 70% of the targeted regions and notably, at least 50X coverage was achieved in 60% of the targeted TRs. Longer sequence reads and optimized read alignment methods facilitated effective analysis of TRs and these findings were independently validated by fragment analysis method. The target enrichment based sequencing of TRs provides new opportunities to explore the impact of genetic variation on complex diseases.

POSTER 7

Mr Naga Kasinadhuni, Mr Lavinia Gordon

BIOINFORMATICS APPROACHES, CHALLENGES AND PERFORMANCE OF DIVERSITY PROFILING ON THREE NEXT GENERATION SEQUENCING PLATFORMS

ABSTRACT

High-throughput sequencing is reshaping the landscape of microbial studies, revealing the hidden secrets of the "uncultured microbial world". Diversity profiling has enhanced the discovery and profiling of microbial species that cannot be isolated into a pure culture, by sequencing molecular markers such as the conserved and hypervariable regions of ribosomal RNA. Most samples consist of an extremely heterogeneous microbial community and may contain thousands of species that vary significantly in proportion from one species to the next. Sequencing these samples using high-throughput platforms outputs an enormous amount of data, ranging from 10^5 to 10^9 reads per run, with each platform producing different read lengths, quality, error rates and data formats. Despite this abundance of data, deciphering some meaningful information presents a serious challenge to bioinformaticians. From QC to classifying the sequenced amplicons, a series of various complex computation methods are required. The Australian Genome Research Facility (AGRF) currently operates a microbial diversity profiling service targeting various regions of 16S, ITS and 18S microbial genes. To compare and evaluate the accuracy and reproducibility of various platforms, AGRF pooled both artificial control samples and environmental samples, and sequences using GSFLX, MiSeq and Ion Torrent NGS platforms. The resulting data was analysed and classified at different taxonomic levels using various bioinformatics pipelines. Here we present the pipelines and results comparing the three NGS platforms.

POSTER 8

Charity W Law, Katarina Matthes, Malgorzata Nowicka, Yifang Hu, Yang Liao, Gordon K Smyth and Mark D Robinson

THE EXTENSION OF VOOM TO TEST PREFERENTIALLY SPLICED EVENTS IN RNA-SEQ COUNT DATA

ABSTRACT

In RNA-sequencing, methods for the detection of changes in isoform usage can be classified into two broad categories: counting (e.g., exon-level) reads according to existing annotation catalogs, or assembling transcripts and estimating their abundances. Either way, these are used in downstream statistical tests to call changes in isoform preference resulting from alternative splicing. Current methods for differential splicing (DS) analyses can be further improved both statistically and in terms of computational expense.

Unlike other existing methods that follow discrete probability distributions, we present an extension to voom (mean-variance modeling at the observational level) which, after application of precision weights, performs DS analyses under a normal distribution. The method, voom-diffSplice, tests for preferentially spliced transcripts at both the exon- and gene-level. Using both simulated and real data, we show that voom-diffSplice performs favorably against existing methods in terms of error control, it produces biologically relevant results, and only takes a small fraction of the computation time required by other existing methods.

POSTER 9

Mr David Lawrence, Dr Andreas Schreiber, Dr Jinghua Feng, Mr Joel Geoghegan

VARIANTGRID: A GRAPHICAL INTERFACE FOR A SCALABLE VARIANT DATABASE

ABSTRACT

Sequencing and variant calling is an increasingly popular means of detecting individual genetic variation, promising to deliver understanding, diagnosis and treatment of many genetic diseases such as inherited disorders and cancer.

Extracting meaning or a diagnosis from this data has proven difficult, requiring an expert knowledge of phenotype, genes and pathways while simultaneously managing very large datasets. These skills are often not found together, slowing progress. We have built a new user interface specifically designed to allow non-programming experts to search through millions of variants.

Our drag & drop interface allows users to create a Directed Acyclic Graph (DAG) of variant filtering and calculation operations. Nodes change to reflect configuration and results, allowing a high level view of the analysis. These can be connected together to build custom analyses, which are converted into efficient queries and run on large databases.

POSTER 10

Dr Xi Li, Mr Joel Ludbey-Bruhwel, Mr Ondrej Hlinka, Dr Josh Bowden, Dr Tim Ho, Dr Annette McGrath

A COMPARISON STUDY OF CPU AND GPU-BASED SHORT READ ALIGNMENT TOOLS

ABSTRACT

The development of efficient short-read alignment tools has still been considered as a challenging topic in Bioinformatics due to the largely accumulated short sequence reads generated from the high-throughput but low-cost next generation sequencing (NGS) platforms. Since last few years, a number of graphic processing units (GPU)-based short-read aligners have been proposed to fully utilize the parallel power of GPUs (e.g., SOAP3, Barracuda, and PEANUT), which aim to further accelerate the alignment process without the loss of accuracy in contrast to the CPU ones. Especially, NVidia has implemented bowtie (NVbowtie) for GPUs so that it affords the opportunity to make a direct comparison of tool's performance under CPU vs GPU. Here, we have carried out a comparative study to investigate the capabilities of GPUs and GPU-aware short-read aligners against CPU ones to understand in what circumstances it would be beneficial to use GPUs and what limitations there are in our current GPU infrastructure. The main facets to examine are speed, accuracy and limitations in terms of data size. This presentation outlines the datasets and tools we chose, the experimental workflows, preliminary outcomes and related experience we have learnt.

POSTER 11

Dr Webber W.P. Liao, Mr Eric H. Powell, Mr Alexander Varlakov, Mr Gavin D. Graham

BRAEMBL GALAXY CONNECT SUITE: HARNESSING THE DATA AND COMPUTE RESOURCES OF THE POWERFUL BRAEMBL COMPUTER CLUSTER THROUGH THE FRIENDLY GALAXY INTERFACE

ABSTRACT

From detecting sequence contamination to inferring protein homology, BLAST is applied in many aspects of bioscience research. While queries of a single sequence can be performed using the NCBI or EBI websites, queries with hundreds of sequences or more are not so straight-forward. One may use the programmatic access provided by websites or set up a BLAST server. However, both require technical knowledge and are not suitable for biologists who are uncomfortable with programming and configuring systems.

Galaxy is a web-based platform for bioinformatics tools. Its friendly interface allows even a novice user to perform simple tasks. Despite this, the existing plugins for BLAST still require configuring the tool and maintaining the necessary databases, which is not trivial for non-bioinformaticians.

To relieve users from the tedious configuration and maintenance, we have developed a new plugin (BRAEMBL Galaxy Connect) that harnesses the data and compute resources of Bioinformatics Resource EMBL Australia (BRAEMBL) and makes them accessible through Galaxy. Once queries are constructed within Galaxy, they are sent to the BRAEMBL server for processing, leaving the Galaxy server's computing power for other tasks. Results are presented in a standard BLAST report as well as a tabular format for downstream analyses. The underlying JDispatcher system also generates visualisations of the results (when available), which are included in an HTML output for easy interpretation.

The cloud-based architecture of "BRAEMBL Galaxy Connect" offers an intriguing alternative to traditional high-performance computing and shows promise for addressing other computational challenges in the field of bioinformatics.

POSTER 12

Mr Andrew Lonsdale

COMBINE, A GROUP AIMED AT STUDENTS AND EARLY-CAREER RESEARCHERS IN BIOINFORMATICS, COMPUTATIONAL BIOLOGY AND SYSTEMS BIOLOGY

ABSTRACT

COMBINE is a student-run Australian organisation for researchers in computational biology, bioinformatics, and systems biology. COMBINE is the official International Society for Computational Biology (ISCB) Regional Student Group (RSG) for Australia.

We aim to bring together students and early-career researchers from the computational and life sciences for networking, collaboration, and professional development.

Australia has many research institutes, each with their own cohorts of students. Aside from conferences, there are few opportunities that bring these students together, allowing them to discover the different kinds of research going on at other institutes.

COMBINE aims to bridge this institutional divide by organising seminars, workshops and social events, also as the yearly COMBINE Student Symposium. Together, these events allow students to connect with each other and build a network in a casual environment.

POSTER 13

Ms Vesna Lukic, Ms Karen L Oliver, Ms Ashley Marsh, Dr Natalie Thorne, Professor Ingrid E Scheffer, Professor Samuel Berkovic, Ms Kate Pope, Associate Professor Paul J Lockhart, Dr Richard J Leventer, Associate Professor Melanie Bahlo

IN SILICO GENE PRIORITISATION WITH BRAINGEP

ABSTRACT

BrainGEP is an R package that makes use of developing and adult Allen Human Brain Atlas data to perform *in silico* gene prioritisation for neurological diseases. Given a set of reference genes (genes previously identified as containing disease causing variants), BrainGEP will calculate pairwise gene correlations using Pearson's and Spearman's correlation measures using a weighted sum approach with weightings derived from the sample variances of the individual pairwise correlation measures. These correlations are then compared to a set of randomly chosen genes, using the empirical cumulative distribution function. The reference gene correlations can also be used to construct networks and correlation plots as additional ways to gauge the strength of the reference gene set. The reference genes can be used to prioritise a set of candidate genes by calculating the connectivity between each candidate gene and the set of reference genes.

We show results from a recently published paper by Oliver *et al* in which a set of 179 candidate Epileptic Encephalopathy (EE) genes were prioritised according to their connectivity with a set of 29 genes known to cause EE. Two of the highest-ranking genes candidate genes prioritised by BrainGEP have subsequently been confirmed as true EE genes. In a second study of a consanguineous family affected with Agenesis of the Corpus Callosum (AgCC), BrainGEP ranked the gene encoding the causal mutation the highest, following linkage filtering. These results demonstrate the benefit of incorporating BrainGEP in gene discovery efforts, as a proven *in silico* prioritisation mechanism for brain diseases.

POSTER 14

Mr Aaron Lun, Professor Gordon Smyth

CONTROLLING THE ERROR RATE CORRECTLY WHEN INTERSECTING GENE LISTS

ABSTRACT

Intersection of lists of differentially expressed (DE) genes is often performed to identify genes that are changing across multiple comparisons. Knowing that a gene is potentially involved in multiple biological processes can provide more insights into its function. However, as with all statistical procedures, a degree of uncertainty is involved in the intersection operation. This uncertainty is represented by the type I error rate of the intersection, i.e., the probability of obtaining a gene in the intersected subset that is actually non-DE in one or more comparisons.

A common strategy is to identify DE genes in each comparison at a particular false discovery rate (FDR), and then identify the shared subset of genes that are DE in each comparison. This approach will not control the type I error rate, nor will it control the corresponding FDR in the intersected subset. Proper control of the type I error rate should instead be maintained with an intersection-union test (IUT). This is a very conservative test in most cases, and will result in loss of detection power.

Here, the inappropriateness of FDR-based intersections and the need for the IUT are demonstrated with some examples. A more powerful version of the IUT is also proposed for cases involving two DE comparisons that are technically independent, i.e., using separate datasets. The performance of this new method is tested on both simulated and real data.

POSTER 15

Paul Maclean, Vanessa Cave, Marlon Reis and Christine Couldrey

INTEGRATING NEXT GENERATION SEQUENCING DATA WITH ADDITIONAL 'OMICS DATASETS

ABSTRACT

It is becoming more and more common to collect data from multiple 'omics sources on an individual sample. Exploration of 'omics data can shed light on the molecular factors that help determine the phenotype of an individual, and a greater understanding of the underlying biological system can be achieved by simultaneously looking at data from multiple sources.

This talk provides a brief overview of integrating Next Generation Sequencing data with additional 'omics datasets, and highlights some methodology and considerations for a new biology.

The talk is centred on research on livestock at AgResearch Ltd.

POSTER 16

Dr Yalchin Oytam, Dr Fariborz Sobhanmanesh, Dr Jason Ross, Dr Konsta Duesing

RISK-CONSCIOUS CORRECTION OF BATCH EFFECTS: MAXIMISING INFORMATION EXTRACTION FROM HIGH-THROUGHPUT GENOMIC DATASETS

ABSTRACT

Identification and effective removal of batch effects is a major challenge in genomic datasets. Batch effects are structured technical noise which persist under best laboratory practice, and typically account for more of the data variance than any of the experimental factors. Even with ideal experimental design, they artificially inflate within-group variances, significantly reducing the power of statistical tests, and resulting in actual effects going undetected. Conversely, if "overcorrected" – i.e. genuine biological variance is removed along with batch noise – the result is an artificial deflation of within-group variances and as a consequence, false positives. Here, we present a novel technique which maximises the removal of batch effects, with the constraint that the probability of overcorrection is kept to a fraction which is set by the end-user.

We benchmark the new technique against the leading technique currently used, using data from published studies, with the two performance criteria being removal of (batch) noise using guided-PCA as measure (Reese et. al, 2013), and preservation of (biological) signal. The new method performs better than the leading method on both fronts simultaneously. For all datasets, while both methods appear to remove batch noise (guided-PCA p-values >> .05) our method does so with markedly lower probabilities for batch effect persistence, and manages to do so while removing significantly less variance from the datasets. Evidence suggests that the novel method for batch correction, applicable to microarray and sequencing data alike, is a notable advance on techniques currently used. A software package has been developed for public release.

POSTER 17

Dilmi Perera, Diego Chacon, Julie A. I. Thoms, Rebecca C. Poulos, Dominik Beck, Peter J. Campbell, John E. Pimanda and Jason W. H. Wong

ONCOCIS : AN ANNOTATION TOOL FOR CIS-REGULATORY MUTATIONS IN CANCER

ABSTRACT

Whole genomes are being sequenced at an accelerated pace but annotation of mutations and inference of their functional significance remains challenging. Whilst a myriad of tools are available for the annotation of protein coding mutations, few are suited for annotating non-coding mutations. To this end, we have developed, OncoCis, to provide an easy to use web service for researchers to annotate and prioritize potential causal cis-regulatory mutations. In order to annotate non-coding mutations, OncoCis integrates publicly available datasets from genome-wide DNase-seq and histone ChIP-seq data obtained from ENCODE and the Epigenome Atlas, to identify mutations that occur within potential *cis*-regulatory regions in a cell type specific manner. These mutations are further annotated with sequence conservation scores and searched for possible removal or creation of transcription factor consensus binding motifs. Finally, FANTOM 5 data combined with the GREAT tool is used to map mutations to the most likely gene that it may be regulating. If gene expression data is available, a fold change is calculated for each of the mapped genes. We have applied this method to whole genome sequencing data from 21 breast cancer samples. Eighteen putative *cis*-regulatory mutations were identified in a number of samples that significantly increased the expression of the mapped gene when compared with other samples without the mutation.This tool will be invaluable to researchers seeking to prioritise large lists of mutations for evaluating their functional relevance.transcription factor consensus binding motifs. Finally, FANTOM 5 data

combined with the GREAT tool is used to map mutations to the most likely gene that it may be regulating. If gene expression data is available, a fold change is calculated for each of the mapped genes. We have applied this method to whole genome sequencing data from 21 breast cancer samples. Eighteen putative *cis*-regulatory mutations were identified in a number of samples that significantly increased the expression of the mapped gene when compared with other samples without the mutation.This tool will be invaluable to researchers seeking to prioritise large lists of mutations for evaluating their functional relevance.

POSTER 18

Dr Belinda Phipson, Dr Alicia Oshlack

DIFFVAR: A NEW METHOD TO TEST FOR DIFFERENTIAL VARIABILITY APPLIED TO DNA METHYLATION IN CANCER AND AGING

ABSTRACT

DNA methylation is crucial for normal development, however methylation changes are known to accumulate with age; and aberrant methylation patterning is associated with many diseases. During both tumour development and aging, a global loss of genome wide DNA methylation combined with gains in CpG island promoter methylation has been observed. To date, the main focus when analysing DNA methylation data has been on detecting differences in the mean levels of methylation between sample groups. Recently, it has been speculated that identifying features that differ in terms of variability may be just as relevant for understanding disease phenotypes. Methods for detecting differential variability as opposed to differential methylation have not been well addressed. Our new method, DiffVar, is inspired by Levene's test and based on an empirical Bayes hierarchical framework. DiffVar is computationally efficient while being robust against non-normality and outliers. The linear modelling framework allows any experimental design to be accommodated. We applied DiffVar to several publicly available cancer datasets from The Cancer Genome Atlas, as well as a publicly available aging dataset, all of which had samples profiled using Illumina's popular 450K human methylation array. Unsurprisingly, the majority (>95%) of the significant differentially variable CpG sites in all cancer datasets tested showed more variability in the cancer samples compared to the normal samples. Interestingly, in the aging dataset, 97% of the significant differentially variable CpGs were more variable in centenarians than in newborns. A closer look at the genomic composition of the significant CpGs revealed that a greater proportion of the top ranked differentially variable CpGs were found in islands compared to differentially methylated CpGs, a phenomenon that was consistently observed across all datasets. DiffVar is available in the Bioconductor R package *missMethyl*.

POSTER 19

Dr Ashfaqur Rahman, Dr Andrew Hellicar, Dr Daniel Smith, Dr John Henshall

ALLELE FREQUENCY CALIBRATION IN POOLED DNA SAMPLES: A MACHINE LEARNING APPROACH

ABSTRACT

We present a machine learning method to estimate allele frequency for SNP based genotyping of DNA pools. DNA pooling is a method commonly used to reduce the costs of genome wide association studies. DNA samples of multiple individuals are combined into pools and genotyped. The raw array output is directly used to compute allele frequency of each SNP.

The raw array output from pooled DNA samples is subject to greater genotyping inaccuracies than individual samples and it is required to adjust the allele frequency estimates. We present a supervised machine learning approach to translate raw array output to allele frequency. Given the genotypes of individuals that constitute the pool, we compute the true allele frequency and train a machine learning method to produce a mapping between raw array outputs to the true allele frequency. We then use this map to

genotype pools with no prior of the individuals constituting the pools.

Two experiments were conducted each using the same Sequenom iPLEX panel which generated results for 61 SNPs. The first larger experiment genotyped 1041 individuals along with 22 pools samples. The pool sizes varied between 18 to 26 individuals all drawn from the 1041 individuals. A second experiment genotyped a sub set of 78 individuals from the first experiment. We compared the performance of the proposed calibration method with some of the existing results. The RMSE error obtained using no calibration, commonly used piecewise linear transformation, and the proposed local–global learner fusion method are 0.135, 0.120, and 0.080 respectively.

POSTER 20

Mr Scott Ritchie, Dr Peter Würtz , Ms Artika Nath, Dr Gad Abraham, Dr Antti Kangas , Dr Aki Havulinna, Dr Pasi Soininen, Dr Kristiina Aalto, Dr Johannes Kettunnen, Dr Michael Inouye

INTEGRATIVE OMICS ANALYSIS ELUCIDATES ROLE FOR INFLAMMATION AND NEUTROPHIL DEGRANULATION IN GLYCOPROTEIN-ASSOCIATED ALL-CAUSE MORTALITY RISK

ABSTRACT

Integration of molecular information captured from high-throughput platforms has shown immense promise for the elucidation of novel biological processes underlying complex disease phenotypes. Here, we integrate genomic, transcriptomic, metabolomic, and cytokine data in two Finnish population cohorts to identify biological processes underlying elevated circulating alpha-1-acid glycoprotein (AGP) levels, a recently discovered biomarker for 5-year all-cause mortality risk. Circulating AGP is influenced by a range of physiological effects; for example AGP increases in response to inflammation, infection, or tissue-injury, and interacts with a variety of drugs. Identifying and understanding the precise biological processes underlying AGP is crucial for determining clinical strategies for intervention and prevention in at-risk individuals. Analysis of a cytokine panel in a longitudinal Finnish cohort (N = 2200) showed an association between low-grade chronic inflammation with (future) elevated AGP. Next, we identified AGP-associated transcriptional networks through weighted gene coexpression network analysis on whole blood samples. Four transcriptional networks were found to be significantly associated with AGP, and these sub-networks and their associations replicated in an independent cohort. Adjustment for lipid associated variation identified one sub-network, enriched for host response to bacterial and fungal infection. Further evaluation revealed that its core genes code for products released during neutrophil degranulation, and Mendelian Randomisation was used to infer causal or reactive relationships between this sub-network and AGP. Our study provides a general framework for integrative analysis of multi-omic data sources, identifies molecular targets for follow-up studies, and suggests prolonged innate immune response may be a driver of increased mortality risk.

POSTER 21

Matthew E. Ritchie, Ruijie Liu, Gordon K. Smyth

WHY WEIGHT? VARIANCE MODELLING FOR DESIGNED RNA-SEQ EXPERIMENTS

ABSTRACT

Outlier samples are relatively common in RNA-seq experiments and the root cause of such variation is generally unknown. In small experiments, the analyst is left with the difficult decision of what to do: removing the offending sample may reduce variation, but at a cost of reducing power, which can limit our ability to detect biologically meaningful changes. A compromise is to use all of the available data, but to down-weight the observations from the outlier sample in the analysis. In this presentation we describe a statistical approach that allows this by modelling heterogeneity at both the sample and observational level in the differential expression analysis. Using both simulations and real data, we tease apart scenarios where this strategy leads to a more powerful analysis from those when it is better to discard the outlier altogether. Our approach is implemented in the open-source limma package available from Bioconductor (http://www.bioconductor.org).

POSTER 22

Alan F. Rubin, Lea M. Starita, Stanley Fields, Terence P. Speed, and Douglas M. Fowler

STATISTICAL METHODS AND SOFTWARE FOR FUNCTIONALLY CHARACTERIZING EVERY SINGLE MUTATION IN YOUR FAVORITE GENE

ABSTRACT

The use of high-throughput DNA sequencing has rapidly expanded catalogues of normal and disease-associated variation, but the functional consequences of most mutations are unknown. In deep mutational scanning, selection for protein function applied to a library of protein variants is combined with high-throughput DNA sequencing to directly measure the activity of hundreds of thousands of variants of the protein easily and cheaply. This approach helps to bridge the gap between variant identification and interpretation, allowing researchers to elucidate sequence-function relationships at high resolution. The resulting data can be used in a variety of contexts, from aiding the assessment of clinical variants to guiding protein engineering. Despite the growing popularity of deep mutational scanning, there are few tools and no formal statistical methods available to help analyze these complex datasets. Here we present a novel method for assigning functional scores and statistical significance to all variants in a deep mutational scanning dataset based on weighted regression. Our methods are implemented as part of Enrich 2, a software package that makes the initial data analysis accessible to experimental biologists while providing an extensible framework for bioinformaticians manipulating these large datasets. We share our results from applying these methods to deep mutational scans of BRCA1, the WW protein-binding domain of YAP65, and other targets. Enrich 2 fully supports barcode sequencing, complex experimental designs involving controls and biological replicates, and deep mutational scans of noncoding sequences, such as structural RNAs or regulatory elements.

POSTER 23

Mr Rick Tankard, Dr Carol Dobson-Stone, Professor Martin Delatycki, Associate Professor David J Amor , Dr Katherine B Howell, Professor Elsdon Storey, Dr Richard J Leventer, Associate Professor John Kwok, Associate Professor Paul J Lockhart, Associate Professor Melanie Bahlo

A STATISTICAL METHOD TO DETECT PATHOGENIC SHORT-TANDEM REPEAT EXPANSIONS IN MASSIVELY-PARALLEL SEQUENCING

ABSTRACT

Expansions of short-tandem repeats (STR) in humans are responsible for over twenty neurological Mendelian disorders including spinocerebellar ataxias, intellectual disabilities, epilepsies and Huntington's disease. Current methods of analysing massively parallel sequencing (MPS) for STR size are focused on genotyping alleles that are smaller than read-fragment size, whereas pathogenic STR expansions can be larger than the fragment size, thus methods are needed to detect pathogenic sized STRs.

We are building a statistical classifier to determine, in an individual with a potential repeat expansion disorder sequenced with MPS, which STR loci have expansions larger than those observed in a healthy population. Here we explore a chi-squared statistic based on binning of read alignments over STR loci. As the binning is likely to break independence assumptions required for a Pearson's chi-squared test, we perform permutation testing to derive the distribution of the chi-squared statistic and generate meaningful p-values. As MPS data of samples with known pathogenic expansions is difficult to obtain, we instead assess the performance with 400 microsatellite markers (ABI PRISM Linkage Mapping Set v2.5) of thirty-two samples, four of which have Illumina whole-genome sequencing (WGS). The largest microsatellite marker allele across the thirty-two samples serve as 'expansions', with those largest in an MPS sample used to assess the power and type I error of our classifier to detect smaller expansions.

POSTER 24

Dr Paul Wang, Dr Wendy Parker, Dr David Yeung, Professor Susan Branford, Dr Andreas Schreiber

OPTIMISING MULTI-CALLER APPROACHES TO SOMATIC VARIANT CALLING

ABSTRACT

Cancer research has benefited greatly from recent improvements in next-generation sequencing methods that have led to significantly increased sequencing depth at relatively low cost. In particular, detection of somatic variants, which often require matched normal-tumour samples and greater detection sensitivity (due to low tumour allele fraction) than germline variants, has become technically and economically feasible.

Several dedicated somatic variant-calling algorithms have been developed recently. Studies that examine these variant callers show that there is a much greater discrepancy between somatic variant callers than germline variant callers. In most cases, the recommended approach is to use multiple callers and rely on their overlap to gauge confidence for variant calling.

Using data from tumour and normal samples of chronic myeloid leukaemia patients, we found that while variants with high caller consensus are much more likely to be true variants (in agreement with other studies), there exist many validated variants that are likely biologically relevant but have low caller consensus.

With the aim to achieve better understanding of the differences between caller behaviour, we carried out in-depth analysis of eight currently available somatic variant callers (MuTect, SomaticSniper, Seurat, Shimmer, EBCall, Strelka, Virmid, and VarScan), and also examined the impact of various processing steps, such as duplicate read removal and indel realignment. We will report these findings and also make recommendations on methods that can improve the accuracy and confidence of somatic variant calling.

POSTER 25

Dr Katrina Bell, Dr Alicia Oshlack

DIFFERENTIAL EXON USAGE DETECTION WITH RNA-SEQ USING ALTERED COUNT SUMMARISATION OPTIONS

ABSTRACT

RNA-seq is an invaluable tool for detecting both differential gene expression and alternative isoform use between experimental conditions, tissues or cells. Alternative splicing is a powerful form of gene regulation that has received less attention than gene differential expression in the past. Currently there are two main approaches for detecting differential isoform expression using RNA-Seq data. One requires the reconstruction of each transcript with subsequent testing for differential expression between samples (e.g. Cufflinks followed by CuffDiff). The second method utilizes a read-count approach at the exon level in order to determine genes with differential exon usage between samples (e.g. Bioconductor; DEXseq, EdgeR-spliceVariants and Voom -diffSplice).

The differential exon usage (DEU) approach taken by the DEXseq method has several advantages. Firstly, the difficult task of reconstructing transcripts and estimating expression levels is not required for testing. Secondly, the count based approach in DEXseq builds on well-established rigorous statistical testing techniques developed for RNA-seq analysis. However, in order for testing for differential DEU, DEXseq requires the RNA-seq reads to be summarised into a count matrix containing the number of reads overlapping known exons. An RNA-seq read has the potential to overlap multiple exons and as sequencing read length is increasing, this is becoming a more common occurrence. It is not clear however how these reads should be treated. One approach is to assign a single read to each overlapping exon and count it multiple times. Other options include only counting the read in the first exon or proportional counting where a read is only counted once per experiment, with the count divided equally among all overlapping exons. Here we investigate how a DEXseq analysis is affected by modifying the count summerisation process to allow only one count per read compared to multiple exon counts per read.

POSTER 26

Leah Alcock, Scott Kuersten, Jim Pease, Agata Czyz, Anupama Khanna, Dixie Gabel, Victor Ruotti, Fraz, Syed, Ramesh Vaidyanathan

WHOLE-GENOME BISULFITE SEQUENCING (WGBS) AN IMPROVED "POST-BISULFITE" CONVERSION LIBRARY CONSTRUCTION METHOD FROM LOW GDNA INPUTS

ABSTRACT

Epicentres continued efforts to innovate, has led to the development of a novel library construction method called EpiGnome™ Methyl-Seq, for preparing sequencing libraries from bisulfite-converted genomic DNA. Current whole-genome bisulfite sequencing methods require substantial amounts of starting DNA to compensate for the loss due to bisulfite-mediated DNA degradation. Additionally, conventional methods, involve laborious DNA shearing, ligation of methylated sequencing adaptors and bisulfite conversion of unmethylated cytidine nucleotides. The main challenge is that approx. 90% of the adaptor tagged DNA is degraded during the bisulfite conversion step. As a consequence, a majority of DNA fragments contained in the adaptor tagged sequence library lose at least one flanking adaptor rendering them incapable of being sequenced. To address and overcome these issues, the EpiGnome™ Methyl-Seq kit "post-bisulfite" method features bisulfite conversion of cytosines prior to addition of adaptor sequences required for cluster generation and sequencing. Thus DNA fragments, independent of size and position of bisulfite mediated strand breaks, remain as viable templates for library construction and sequencing. This novel approach exhibits increased sensitivity and efficiency, along with improved coverage for detecting methyl-cytidine nucleotides.

POSTER 27

Associate Professor Jeffrey Craig, Dr Mark Cruickshank, Dr Christiane Theda, Professor Peter Davis [3], Dr Penny Sheehan, Dr David Martino, Dr Alicia Oshlack, Ms Yun Dai, Dr Richard Saffery , Professor Lex Doyle

EPIGENETIC CHANGES IN SURVIVORS OF PRETERM BIRTH: EFFECT OF GESTATIONAL AGE AND EVIDENCE FOR A LONG TERM LEGACY

ABSTRACT

Preterm birth confers a high risk of adverse long-term health outcomes for survivors, yet the underlying molecular mechanisms are unclear. We hypothesized that effects of preterm birth can be mediated through measurable epigenetic changes throughout development. We therefore used a longitudinal birth cohort to measure the epigenetic mark of DNA methylation at birth and 18 years comparing survivors of extremely preterm birth with infants born at term.

Using 12 extreme preterm birth cases and 12 matched, term controls, we extracted DNA from archived neonatal blood spots and blood collected in a similar way at 18 years of age. DNA methylation was measured at 347,789 autosomal locations throughout the genome using Infinium HM450 arrays. Representative methylation differences were confirmed by Sequenom MassArray EpiTyping.

At birth we found 1,555 sites with significant differences in methylation between term and preterm babies. At 18 years of age, these differences had largely resolved, suggesting that DNA methylation differences at birth are mainly driven by factors relating to gestational age, such as cell maturity and to some extent, cell composition. Using matched longitudinal samples, we found suggestive evidence for an epigenetic legacy of preterm birth, identifying persistent methylation differences at several genomic loci. Longitudinal comparisons uncovered a significant overlap between sites that were differentially-methylated at birth and those that changed with age. We found an unexpectedly high number of birth- and age-differentially-methylated probes (314) showing a higher level of DNA methylation in term infant's blood compared to both preterm infants' and adult blood, suggesting that developmentally-associated epigenetic trajectories may not always be linear.

We present evidence for widespread methylation differences between extreme preterm and term infants at birth that are largely resolved by 18 years of age. These results are consistent with methylation changes associated with blood cell development, immune induction, cellular composition and age at these time points. Finally, using this small sample size we identified a number of sites that were persistently altered in the preterm group and may be associated with a long term epigenetic legacy of preterm birth.

POSTER 28

Dr Christoffer Flensburg, Dr Alicia Oshlack

A COMPARISON OF CONTROL SAMPLES FOR CHIP-SEQ OF HISTONE MODIFICATIONS

ABSTRACT

The advent of high-throughput sequencing has allowed genome wide profiling of histone modifications by Chromatin ImmunoPrecipitation (ChIP) followed by sequencing (ChIP-seq). In this assay the histone mark of interest is enriched through a chromatin pull-down assay using an antibody for the mark. Due to imperfect antibodies and other factors, many of the sequenced fragments do not originate from the histone mark of interest, and are referred to as background reads. Background reads are not uniformly distributed and therefore control samples are usually used to estimate the background distribution at any given genomic position. The Encyclopedia of DNA Elements (ENCODE) Consortium guidelines suggest sequencing a whole cell extract (WCE, or "input") sample, or a mock ChIP reaction such as an IgG control, as a background sample. However, for a histone modification ChIP-seq investigation it is also possible to use a Histone H3 (H3) pull-down to map the underlying distribution of histones.

Here we generated data from a hematopoietic stem and progenitor cell population isolated from mouse foetal liver to compare WCE and H3 ChIP-seq as control samples. The quality of the control samples is estimated by a comparison to pull-downs of histone modifications and to expression data. We find minor differences between WCE and H3 ChIP-seq, such as coverage in mitochondria and behaviour close to transcription start sites. Where the two controls differ, the H3 pull-down is generally more similar to the ChIP-seq of histone modifications. However, the differences between H3 and WCE have a negligible impact on the quality of a standard analysis.

POSTER 29

Dr Aliaksei Z. Holik, Laura A. Galvis Vargas, Julie Pasquet, Aaron T. L. Lun , Dr Marnie E. Blewit, Dr Matthew E. Ritchie, Dr Marie-Liesse Asselin-Labat

EZH2 IS ESSENTIAL FOR LUNG DEVELOPMENT, MAINTENANCE OF EPITHELIAL CELL IDENTITY AND CELL FATE DETERMINATION

ABSTRACT

The role of epigenetic regulation in developmental processes has become increasingly evident. The polycomb repressive complex 2 (PRC2) has been shown to play a fundamental role in controlling stem cell function in a number of organs. However, its involvement in lung development, as well as the role of epigenetic regulation in general, remains poorly understood. We explored the role of the catalytic subunit of PRC2, Ezh2, during lung morphogenesis. Loss of Ezh2 in the lung epithelium led to defective lung formation and perinatal mortality. We show that Ezh2 is critical for airway lineage specification and alveolarization. We use gene expression profiling of embryonic lung mesenchyme and epithelium deficient in Ezh2, combined with ChIP-seq analysis of H3K27 tri-methylation, as well as publicly available gene expression and histone modification datasets to elucidate potential mechanisms of Ezh2 function in embryonic lung development.

POSTER 30

Miss Siobhan Hughes

GENOMEWIDE ANALYSIS OF PROTEINS THAT BIND TO DNA AND REGULATE GENE EXPRESSION

ABSTRACT

Regulation of gene expression is a complicated process, subject to different mechanisms operating at different levels. At the genomewide level, chromatin immunoprecipitation and next-generation sequencing are being used to interrogate the coincident and allele-specific binding of the proteins CTCF, Cohesin, ATRX and MeCP2 in the murine brain, a tissue that expresses a significant proportion of imprinted transcripts. Identifying regions co-binding these proteins will help generate a model to understand how these proteins influence gene expression, particularly at imprinted loci. Using these tools we aim to understand more about the proteins that participate in the genomic landscape around imprinted loci and drive this unusual mode of gene regulation. These loci will act as models for studying DNA binding proteins and their roles in transcription.

At the level of the individual locus, mechanisms of gene regulation are being investigated using imprinted retrogenes as a model. Retrogenes are transcriptionally active intronless genes located within an intron of a 'host' gene. The role of epigenetic factors influencing gene transcription are being investigated at the *Mcts2/H13* locus. Expression of an intronic retrogene can cause premature termination of a 'host' transcript from the same allele. Our hypothesis is that this premature termination is caused by transcription of the retrogene interfering with host gene transcription. We have designed a construct based on this locus, which will allow us to regulate the expression through the retrogene/internal promoter to study this in more detail. These studies will provide a mechanistic component to our whole genome analyses.

POSTER 31

Dr Jovana Maksimovic, Dr Yuxia Zhang, Mr Gaetano Naselli, Junyan Qian, Dr Michael Chopin, Dr Marnie E. Blewitt , Prof. Leonard C. Harrison, Dr Alicia Oshlack

EPIGENETICS, T-CELLS AND AUTOIMMUNITY: A BIOINFORMATICS STORY

ABSTRACT

Autoimmune diseases result in an inappropriate immune response against healthy cells. They are likely caused by a combination of genetic predisposition and environmental triggers. Epigenetic phenomena are thought to be the link between the environment, genetic background and development of autoimmune disease.

Regulatory T cells (Treg) are required to prevent the emergence of autoimmune disease and evidence suggests that epigenetic regulation is required to establish their ultimate phenotype and function. To explore the methylation landscape of Treg we analysed genome-wide methylation in human naïve Treg (rTreg) and conventional naïve CD4+ T cells (Naïve) using Illumina 450k arrays. By taking into account the differences between individuals and performing both probe-wise and regional differential methylation analysis we detected 2,315 differentially methylated CpGs between these two cell types, many of which clustered into 127 regions of differential methylation (RDMs). Further analysis showed that the RDMs were enriched for putative FOXP3 binding motifs. Bioinformatic analysis of publicly available data also revealed that CpGs within known FOXP3-binding regions were hypomethylated and that there was an inverse relationship between promoter RDM methylation and gene expression. Thus, hypothesising that methylation limits access of FOXP3 to its DNA targets we proceeded to show that increased expression of the immune suppressive receptor TIGIT was associated with hypomethylation and FOXP3 binding at the TIGIT locus.

We are currently building on this research to develop methylation biomarkers that can be used to estimate the proportion of nTreg in human blood, which may have diagnostic and prognostic applications for autoimmune disease.

POSTER 32

Michelle French, Anne O'Connell, Sara Edwards, Avijit Haldar, Phil Farquhar, Ken Dodds, Sue Galloway, Rudiger Brauning, Peter Johnstone, George Davis, Jenny Juengel

MUTATIONS IN LEPR GENE ARE ASSOCIATED WITH REPRODUCTIVE TRAITS IN DAVISDALE SHEEP

ABSTRACT

Whole genome sequencing of four key sires from the highly prolific Davisdale sheep line were used to identify mutations in candidate genes known to be involved in the regulation of reproductive performance. This identified three single nucleotide polymorphisms (SNPs) that change the amino acid sequence of the Leptin Receptor (*LEPR*) gene. A sequenom assay was designed and we genotyped 750 ewes with records for ovulation rate, first service conception rate and partial failure of multiple ovulations; 600 of these ewes also had records for puberty age. The aim of this study was to determine if the SNPs in the LEPR gene were associated with reproductive traits. Ewes homozygous for the *LEPR* gene mutations either failed to undergo puberty during their first year, or those that did attain puberty were 18 days older than ewes which did not have the *LEPR* mutations. Age at puberty was intermediate between the two homozygous groups in ewes heterozygous for the mutations. Ewes that were homozygous for mutations in the *LEPR* gene had a 15% decrease in ovulation rate ($P<0.001$), and a 12% decrease in their first service conception rate ($P<0.01$). Ewes that ovulated three ova had increased partial failure of multiple ovulations ($P<0.01$), and had 0.2 fewer lambs at mid-pregnancy and at birth ($P<0.01$) compared to ewes that are wild type. Our studies have shown that poorer reproductive performance in the Davisdale line is strongly associated with homozygosity for mutations in *LEPR*.

POSTER 33

Mr Amali Thrimawithana, Mr Bernd Steinwender, Mr Ross Crowhurst, Mr Richard Newcomb

EXPLORING THE PHEROMONE GLANDS OF NATIVE NEW ZEALAND LEAFROLLER MOTHS USING DE NOVO TRANSCRIPTS PRELIMINARY INSIGHTS

ABSTRACT

Many of the enzymes involved in pheromone biosynthesis pathways in moths are yet to be discovered. We have used a transcriptomics-based approach to help to gain insights into the pheromone biosynthesis pathway, with a longer-term goal of understanding the involvement of these genes within this pathway in chemical signalling. We explored the pheromone glands of New Zealand leafroller moths. New Zealand leafroller moths include species from the genera *Ctenopseustis* and *Planotortrix*. We have focused our research on four species (*C. obliquana, C. herana, P. octo and P. excessana*). These species pose a significant threat to important horticultural crops in New Zealand because of their highly polyphagous nature. The sibling species are similar in appearance, but are known to be reproductively isolated in the wild because of distinct sex pheromones. Therefore exploration of the pheromone biosynthesis pathways in these species will be beneficial in understanding the differences among these species. To achieve this we sampled the transcriptomes of isolated pheromone glands from each species and have *de novo* assembled transcripts using a multi-assembler approach. We highlight and discuss the preliminary insights we gained from this multi-assembler approach.

POSTER 34

Dr Christy Vander Jagt, Dr Amanda Chamberlain, Professor Mike Goddard, Dr Leah Marett, Ms Lakshmi Krishnan, Professor Ben Hayes

A BOVINE GENE EXPRESSION ATLAS

ABSTRACT

The subset of expressed genes within a cell convey the unique properties of different cell- and tissue-types. In the cow, the structure of this 'expression space' is largely unknown. Next-generation sequencing allows comprehensive transcriptome surveys to be performed, providing insight into tissue-specific gene expression. In this study we employed RNA-sequencing (RNA-seq) to construct a bovine gene expression atlas, generating a snapshot of gene expression at a single time point across 18 different tissue types in a lactating dairy cow.

100 base paired end RNA-seq reads were generated on an Illumina HiSeq2000 from each tissue (in triplicate). Tophat2 was used to map reads to the Ensembl annotation of UMD3.1 bovine genome assembly. On average 92% of reads aligned per sample (>70% uniquely mapped). Differential expression analysis using DESeq and EdgeR identified tissue-specific transcriptional trends and demonstrated that of the 18 tissues, the two brain tissues had the greatest number of significantly differentially expressed genes (when compared to average gene expression). The relationship between tissues was investigated via hierarchical clustering. Tissues clustered together reflecting their biological relationship (i.e. all muscle tissues clustered together), providing some validation of results. Functional annotation of differentially expressed genes further validated results, with many biological processes and pathways identified as being significantly enriched ($p<0.01$) already having an established role in those particular tissues (i.e. the top most significant Gene Ontology term associated with differentially expressed genes in both white skin and black skin was 'epidermis development').

The bovine gene expression atlas generated by this study will advance our understanding of the dynamics and complexity of bovine transcription and provide a valuable resource for future research on functional genomics in the bovine species.

POSTER 35

Mr Bjorn Dueholm, Dr Bjorn Hamberger, Professor Philip Weinstein, Dr Susan Semple, Professor Birger Lindberg Moller

COMPARATIVE TRANSCRIPTOMICS OF ACACIA ACUMINATA PARASITISED VS. NON-PARASITISED BY THE HEMIPARASITIC SHRUB SANTALUM SPICATUM

ABSTRACT

Acacia acuminata (Raspberry jam tree) is common to Western Australia and often found parasitised by the hemiparasitic shrub *Santalum spicatum* (Australian sandalwood). Australian sandalwood takes up water, nutrients, and carbon from other plants through specialised organs on the roots called haustoria. This parasitism puts the acacia host under some level of biotic stress. Studies have shown a variety of different physiological stress-responses from the plant parasitism in the acacia host; however, how the host responds on a transcriptional level has not yet been investigated.

RNA-Seq is a powerful method for the investigation of gene expression in non-model organisms and can be used for comparative transcriptomics between two or more types of treatments. Experimental plantations in which Raspberry jam tree grows under the same abiotic conditions either alone or together with Australian sandalwood will be utilized for investigating the differential expression of genes in the acacia host under the two treatments.

Parasitic plants have been found not only to affect the hosts but also indirectly to affect other organisms associated with the hosts. This study is the first to utilize RNA-Seq for studying stress-induction in woody hosts by parasitic plants. In addition to advancing our understanding stress in woody hosts it will increase our understanding of the link between belowground and aboveground events and provide genetic information that can be used for studying natural communities.

POSTER 36

Chol-Hee Jung, Professor Mohan Singh, Professor Prem Bhalla

DYNAMIC ACTIVITIES OF NOVEL TRANSCRIPTS IN LEAF AND SHOOT APICAL MERISTEM OF SOYBEAN DURING FLORAL TRANSITION

ABSTRACT

The analysis of high-throughput sequencing of RNAs is capable of revealing the details of transcriptional landscapes in an unprecedented level. We recently conducted whole genome transcriptome analysis of soybean, one of the major crop plants, to uncover the dynamic transcription and regulation during the early floral initiation process (Wong *et al.*, 2013, *PLoS ONE*) focusing on protein coding genes. Soybean is a paleopolyploid that has been subjected to at least two rounds of whole genome duplication events. We are interested to investigate involvement of alternative splicing, antisense transcripts and intergenic transcripts during floral initiation in soybean after inductive short-day treatment. We identified ~107,000 transcripts from ~56,000 loci in the soybean genome using a total of ~120 million paired-end sequences from leave and shoot apical meristem samples taken at different time-points after short-day treatment (floral-inducing). We observed that the shoot apical meristem utilizes a greater repertoire of splicing variants than leaf and that the use of alternative splicing isoforms changes dynamically depending on developmental stages and/or tissue types. Multiple-isoform loci are enriched for antisense transcripts compared to single-isoform loci. The same sense and antisense transcript pair appears to have a different relationship in different tissues, and the same antisense

transcript may interact differently with different sense transcripts in the same locus. Overall, our study improves the annotation of the soybean genome by identifying numerous novel transcripts and demonstrates the complex regulation of splicing variants with or without an association with antisense transcripts during floral initiation.

POSTER 37

David Kainer, Dr Amanda Padovan, Hamish Webb, Professor William Foley, Dr Carsten Kulheim

TOWARDS GENOMIC SELECTION FOR ESSENTIAL OIL YIELD IN EUCALYPTUS AND MELALEUCA

ABSTRACT

The yield of essential oil in commercially harvested Myrtaceae species (e.g. *Eucalyptus polybractea, Eucalyptus loxophleba* and *Melaleuca alternifolia*) is a complex quantitative trait which is itself composed of multiple quantitative traits such as foliar oil concentration, biomass and adaptability. These traits often show large natural variation and some are highly heritable. Foliar essential oils are mostly comprised of mixes of mono- and sesquiterpene compounds and so recent association studies have investigated well characterised genes of the terpene biosynthetic pathway. Analysis of transcript abundance and allelic diversity has revealed that essential oil yield is likely to be controlled by large numbers of quantitative trait loci (QTLs) that range from a few of medium/large effect to many of small effect, with much of the variation occurring beyond the major genes of the biosynthetic pathway.

Genome-wide association studies (GWAS) are often used to investigate the genetic architecture of complex quantitative traits across the whole genome to explain as much variation as possible. However, in naturally outcrossing forest trees the feasibility of GWAS has been limited by the requirement for very high density genotyped markers due to the short range of linkage disequilibrium (~100 bp). Now, with the cost per marker continuing to drop, we investigate the feasibility of genome-wide association and genomic selection for improving essential oil yield in *E. polybractea*. We compare traditional breeding and marker-assisted selection for essential oil yield, and how the use of whole genome re-sequencing and genomic selection may further improve selection of elite individuals.

POSTER 38

Dr Jafar Sheikh Jabbari, Dr Rust Turakulov, Dr Kirby Siemering, Mr Matthew Tinning, Ms Maria Pazos-Navarro, Dr Daniel Real, Dr Matthew N. Nelson

EFFICIENT GENOTYPING-BY-SEQUENCING ENABLES CONSTRUCTION OF THE FIRST LINKAGE MAP FOR THE NON-MODEL LEGUME FORAGE SPECIES, BITUMINARIA BITUMINOSA VAR. ALBOMARGINATA (TEDERA)

ABSTRACT

Single-nucleotide-polymorphisms (SNPs) identified using genotyping-by-sequencing (GBS) methods are increasingly used in many fields of genetics and population genomics. These methods differ in nucleotide diversity during initial sequencing cycles that affects quality and quantity of reads, number of sampled digestion fragments (tags) and sequencing depth required for confident genotype calls. Here we introduce an improved GBS method based on size-selection of double-digested adapter-ligated DNA which increases sequencing yield and quality by producing high diversity libraries. We applied this technique for genotyping a subset of an F2 population from *Bituminaria bituminosa* var. *albomarginata* (tedera), an emerging drought-tolerant forage legume species. By sequencing a barcoded library prepared from 45 progeny and parents in one lane of HiSeq we obtained over 150 M SE reads which were analysed using Stacks pipeline resulting in 845 high-quality SNP markers scored in all samples. Linkage mapping identified 11 linkage groups, close to the haploid chromosome number of this species (n=10). Analysis of synteny via BLASTn between the genomes of tedera and its close relative soybean using tedera GBS

sequence tags revealed the presence of 264 (31.2%) GBS marker tags with one or more significant (<1e^{-10}) matches in the soybean genome and that marker order was conserved in large chromosomal blocks between the two genomes. This finding, along with application of the method to the remaining progeny, will facilitate the transfer of genomic information between the well-resourced model genome of soybean and the 'orphan' genome of tedera.

POSTER 39

Mnasri Sameh, Saddoud Dabbebi Olfa, Ben Saleh Mohamed, Professeur Ferchichi Ali

MICROSATELLITE MARKER-BASED IDENTIFICATION AND GENETIC RELATIONSHIPS OF MILLENNIUM OLIVE CULTIVARS IN TUNISIA

ABSTRACT

Microsatellite markers were used to characterize the millennium olive cultivars localized in nine different archeological sites in Tunisia. Thirty genotypes were considered for genetic fingerprinting using 10 pairs of microsatellite primers. The number of alleles per locus ranged from 3 to 5, with a mean of 3.7 alleles per primer pair (a total of 37 alleles). The observed heterozygosity ranged from 0.4 to 1, while the expected heterozygosity varied between 0.37 and 0.74. The polymorphism information content values ranged also from 0.37 to 0.74. The mean polymorphism information content value of 0.61 for the SSR loci provided sufficient discriminating ability to evaluate the genetic diversity among the millennium cultivars. The UPGMA cluster analyses using Jaccard's index permitted a segregation of the thirty millennium cultivars in three main groups and revealed that most of the millennium cultivars grouped according morphological parameters of the fruit and the endocarp and no clear clustering trends were observed according to their geographic origin. As a sequel to the present work, new surveys should be made in the archeological sites localized in North and the Center of Tunisia to sample more cultivars and to draw a clearer picture of the diversity of the Tunisian millennium olive germplasm.

POSTER 40

Dr Jeremy Shearman, Ms Duangjai Sangsrakru, Ms Panthita Ruang-areerate, Ms Chutima Sonthirod, Ms Pichahpuk Uthaipaisanwong, Ms Thippawan Yoocha, Ms Supannee Poopear, Dr Kanikar Theerawattanasuk, Dr Somvong Tragoonrung, Dr Sithichoke Tangphatsornruang

ASSEMBLY AND ANALYSIS OF A MALE STERILE RUBBER TREE MITOCHONDRIAL GENOME REVEALS DNA REARRANGEMENT EVENTS AND A NOVEL TRANSCRIPT

ABSTRACT

The rubber tree, Hevea brasiliensis, is an important plant species that is commercially grown to produce latex rubber in many countries. The rubber tree variety BPM 24 exhibits cytoplasmic male sterility, inherited from the variety GT 1. We constructed the rubber tree mitochondrial genome of a cytoplasmic male sterile variety, BPM 24, using 454 sequencing, including 8 kb paired-end libraries, plus Illumina paired-end sequencing. We annotated this mitochondrial genome with the aid of Illumina RNA-seq data and performed comparative analysis. We then compared the sequence of BPM 24 to the contigs of the published rubber tree, variety RRIM 600, and identified a rearrangement that is unique to BPM 24 resulting in a novel transcript containing a portion of atp9. The novel transcript is consistent with changes that cause cytoplasmic male sterility through a slight reduction to ATP production efficiency. The exhaustive nature of the search rules out alternative causes and supports previous findings of novel transcripts causing cytoplasmic male sterility.

POSTER 41

Mr Tim Stuart, Prof. Ryan Lister, Dr Taiji Kawakatsu, Dr Natalie Breakfield, Dr Manuel Valdes, Dr Joseph Nery, Dr Xinwei Han, Prof. Robert Schmitz, Prof. Philip Benfey, Prof. Joseph Ecker

SINGLE CELL-TYPE ANALYSIS REVEALS UNIQUE PATTERNS OF DNA METHYLATION IN THE ROOT MERISTEM

ABSTRACT

In the plant genome, methylation of cytosine bases (DNA methylation) plays a critical role in transcriptional silencing of transposable elements (TEs) to protect genome stability. Multiple pathways exist in *Arabidopsis thaliana* for the establishment and maintenance of DNA methylation, including the RNA-directed DNA methylation (RdDM) pathway utilizing a 24 nucleotide (nt) small RNA (smRNA) signal to direct DNA methylation. Recently, single cell-type analysis of Arabidopsis reproductive cells unveiled widespread changes in DNA methylation patterns and smRNA production in the reproductive companion cells, and that these changes play an important role in maintaining TE silencing in the gametes and developing embryo.

In plants, stem cell niches (meristems) responsible for tissue growth and development exist in the root and shoot. To investigate whether a reconfiguration of DNA methylation marks similar to that of the reproductive cells occurs in meristems, we used an integrated analysis of DNA methylation, mRNA and smRNA transcript abundance in five different cell types of the root meristem. Our analysis revealed that the columella, a group of terminally differentiated, short-lived cells situated directly below the root stem cells, contains levels of DNA methylation far higher than any other Arabidopsis cell or tissue type reported to date. Hypermethylation of the columella genome was coupled with an increase in mRNA transcripts encoding components of the RdDM pathway, as well as elevated levels of 24 nt smRNAs needed for RdDM. These changes were restricted to TEs in the columella genome, and the patterns or levels of DNA methylation in expressed genes did not differ between the cells types studied.

As columella cells are terminally differentiated, mutations induced through TE activity are not inherited by any other cells, and consequently would have a small impact on the overall fitness of the plant. However, the enhancement of TE silencing in the columella observed in this study indicates that TE silencing is particularly important in these cells. One possible explanation for this seeming contradiction is that 24 nt smRNAs generated in the columella are transported to the neighbouring root stem cell to reinforce transcriptional silencing of TEs through RdDM, similar to processes thought to occur between reproductive cells and their non-generative companion cells. This establishes the columella as potential companion cells to the root stem cells.

POSTER 42

Mr Anandhakumar Chandran, Mr Li Yue, Mr Seiichiro Kizaki, Ms Le Han, Dr Ganesh Pandian, Mr Syed Junetha, Mr Shinsuke Sato, Mr Junichi Taniguchi, Dr Toshikazu Bando, Dr Hiroshi Sugiyama

HIGH THROUGH-PUT SEQUENCING GUIDES THE DESIGN OF DNA MINOR GROOVE BINDING SMALL MOLECULE WITH IMPROVED BINDING SPECIFICITY

ABSTRACT

In chemical genomics and molecular biology, a small molecule is an organic compound with low molecular weight that is useful in regulating biological process. The development of small molecules that trigger targeted transcriptional activation could ensure better efficacy and a reduction of long-term side effects. Our results show, an artificial small molecule named SAHA-PIP K containing the histone deacetylase inhibitor SAHA and the sequence-specific DNA minor groove binding hairpin pyrrole-imidazole polyamides (PIPs), could rapidly stimulate the meiotic process regulating PIWI pathway

marker genes in human dermal fibroblasts. This first ever result on a small molecule that significantly induces the endogenous regulation of the conserved germ cell genes in a human somatic cells opens up the panorama of chemical mediated reprogramming of somatic cell to a germ cell. But, whether SAHA-PIP's strong biological effects due to its binding specificity in a broad context or else other is still unclear. To characterize the binding specificity of SAHA-PIPs, we carried out high-throughput sequencing (Bind-n-seq) studies to define the recognition motifs of PIP K (PIP without SAHA) / SAHA-PIP K -DNA binding in a genome sized sequence space. SAHA-PIP K showed, high affinity binding compared with PIP K. This induced us to redesign SAHA-PIP without SAHA and with the moiety in the non-core binding region of the ten-ring hairpin polyamide. The results showed the moiety can enforces the closed binding of PIP.

POSTER 43

Mr Arunkumar Padmanaban, Dr Eva Schmidt, Dr Adam Inche, Dr Ruediger Salowsky

END-TO-END SAMPLE QUALITY CONTROL FOR NEXT GENERATION SEQUENCING LIBRARY PREPARATION AND SURESELECT TARGET ENRICHMENT ON THE AGILENT 2200 TAPESTATION SYSTEM

ABSTRACT

Next generation sequencing (NGS) has been widely adapted to address intriguing biological questions and understand mechanisms at the genome level. New developments in sequencing methods and high-throughput technologies has reduced the timeline for genome sequencing from weeks to days. Even though the cost of NGS has significantly reduced, the upfront sequencing library preparation is still complex and time consuming. Several companies offer library preparation workflows to help reduce complexity. In addition, the targeted enrichment of the end libraries for specific genomic subsets can further drive down cost by increasing sequencing coverage. However, library preparation workflows still need to be carefully controlled and monitored to ensure downstream sequencing success. The Agilent SureSelect Target Enrichment solution offers a complete, validated workflow of sample preparation for target specific sequencing with a range of capture sets including all exons, panels like clinical research exome. Here we present data to demonstrate the Agilent 2200 TapeStation system as an ideal and reliable quality control (QC) platform for sizing and quantification analysis of the NGS libraries. The Genomic DNA ScreenTape assay assess the starting genomic DNA sample for quality and quantity, replacing independent measurement systems with a single test, thereby saving sample and time. The sizing and quantification of the intermediate and final NGS libraries can be reliably carried out using the D1000 ScreenTape assay. Thus, the 2200 TapeStation system allows analysis of the starting genomic sample (DNA), through to final QC of the DNA libraries prior to sequencing.

POSTER 44

Dr Li Zhou, Daniel Catchpoole

EVALUATION OF FLUIDIGM SNPTRACE™ PANEL, WITH A COMPARISON TO DNA QUAL AND BIOANALYZER FRAGMENT ANALYSIS, FOR DNA QUALITY ASSURANCE WITHIN THE TUMOUR BANK PROGRAM

ABSTRACT

Translational research is the key component to apply basic science findings into practical medicine. High quality human biological specimens are the foundation of accurate and reproducible research result and will maximize the productivity of translational research. To assure that the best samples are provided to translational research, quality assurance (QA) programmes are an essential part of biobanks' routine.

To identify the most feasible technology for regular biobank QAP daily practice, two technologies promoted from use in biobanking QAP have been compared, the Fluidigm SNP trace panel (Fluidigm, USA) and DNAqual (Eurobio, France). 93 samples, including 80 DNA samples from 4 donors and 13 DNA samples isolated from bone marrow aspirates (BMA) smeared on slides, were examined using both technologies. BMA slides had been stored at the room temperature and exposed to air for 2-14 years. In addition, to create a range of possible conditions occurring in biobank daily practice, DNA samples from donors were treated with multiple freeze-thaw cycles (0-20), varied snap delay period (0-21 days), radiation, sonication, heat, UV and mixed contamination (MC) respectively. DNA qualities have been verified using Bioanalyzer DNA 7500 Kit (Agilent, USA).

The results showed that the Fluidigm panel could detect 50% MC between samples accurately. The sensitivity of the contamination detection need to be further tested. It also reported the degraded DNA qualitatively instead of quantitatively. By contrast, DNAqual can provide the quantitative DNA quality index although it is not designed for detecting MC at all. The accuracy of the index need to be further validated.

Inaddition, the consistency rate of the Fluidigm panel is 100% within the plate due to its qualitative feature, but it is not optimal to compare the results across the plates, whereas due to its qPCR feature, DNAqual can compare the readings across runs. All these results will be further validated and elucidated.

These results form the basis of a better biospecimen QA solution for the Australian biobanking community. Furthermore, we will discuss the ongoing implementation of these new QA platforms into daily biobank practice and how these will benefit translational research involving genomics.

POSTER 45

Associate Professor Richard Allcock

A BETTER WAY - AMPLISEQ EXOME FOR CLINICAL RESEARCH

ABSTRACT

Many laboratories are actively engaged in implementing large-scale sequencing in clinical and diagnostic settings. There is significant debate about the best approach (gene-specific panels, WES, WGS), each of which has benefits and disadvantages. In addition to accuracy and utility, other practical considerations must be taken into account such as cost, scalability, practicality of implementation and infrastructure requirements.

In many settings, WES may strike the appropriate balance between these factors.Since its development, WES has made incremental advances but for the most part is still based upon hybridisation to biotinylated probes. Recently a completely new way to enrich exomes was developed based on massively parallel PCR (AmpliSeq). In combination with the Ion Torrent Chef and Proton sequencer, the process is rapid, flexible, scalable in a variety of settings and highly reproducible. it is now possible to routinely and economically sequence 100X+ coverage exomes).

Exomes enriched by amplification differ significantly from those enriched by hybridisation. First, amplification is more specific than hybridisation and hence very high on-target rates (>94%) are can be achieved. Second, there is less need to sequence deep into introns, with primers sited closer to the exons of interest. Third, analytical approaches need to be modified as a PCR-based approach produces PCR duplicates (by definition). Finally, the process is extremely rapid, allowing us to go from DNA to final analysed result in as little as 36 hours, which potentially allows application in many areas not traditionally amenable to exome level analysis.I will provide an overview of the latest developments using the AmpliSeq Exome RDY kit and its application to clinical research problems in Western Australia.

Tuesday

14 OCTOBER

POSTER 46

Nigel C Bennett, Maryam Jessri, Andrew J Dalley, Camile S Farah

EXOMIC MOLECULAR SIGNATURES SEGREGATE ORAL EPITHELIAL LESIONS BASED ON DISEASE SEVERITY

ABSTRACT

To determine if the accumulation and genetic location of deleterious mutations within oral epithelial lesions correlates with increasing grades of epithelial pathology, 53 archival formalin fixed paraffin embedded (FFPE) oral epithelial biopsies inclusive of 5 oral squamous cell carcinoma, 31 epithelial dysplasia, and 11 samples with no epithelial abnormality were examined using next generation sequencing. Target-enriched exome libraries were constructed from 500ng-3µg gDNA using the SureSelectXT target enrichment system (Agilent Technologies, CA, USA) for SOLiDTM 5500XL multiplexed sequencing (Life TechnologiesTM, CA, USA). Sequenced exomes were mapped to the hg19 reference human genome and variants called using the GATK software. The numbers of variants in each gene were compared in different groups of samples using Kruskal-Wallis and sPLS-DA. All p-values were corrected for multiple comparisons using Bonferroni's correction. Per sample, a mean of 170,906,808±52,255,566 unique reads were generated from which a mean of 79±7.8% were on target with a mean coverage of 137X. A mean of 23,335±2,554 variants were called per sample, from which synonymous variants (mean=11,339±1,258) were excluded. When tested by Kruskal-Wallis, lesions could be significantly discriminated based on the number of exomic variants (non-synonymous, frameshift and non-sense). Multivariate analysis separated the lesions based on their severity using two components and 10 variables. An increased number of exomic variants in different oral epithelial pathological states is supportive of the

field cancerization theory. Oral epithelial biopsies can be separated based on their severity using the number of mutations in their exome.

POSTER 47

Ms Magdalena Budzinska, Dr Thomas Tu, Dr Fabio Luciani, A Nicholas Shackel

GENOMIC EVOLUTION OF HEPATOCYTES TOWARDS HEPATOCELLULAR CARCINOMA

ABSTRACT

Hepatocellular Carcinoma (HCC) is associated with accumulation of number of genetic abnormalities. The vast majority of those events, called passenger mutations, are non-recurrent and dispersed across the cancer genome. A previous study has shown that deleterious passenger mutations (DPM) occur more frequently in tumour tissue compared to surrounding non-tumour tissue. Therefore, we hypothesise that there is a stepwise accumulation of DPM in HCC progression. To address this, we aimed to characterise DPM in whole exome sequencing (WES) datasets from patients in various stages liver disease.

We analysed 3 publically-available WES datasets of HCC and paired non-tumour tissue and generated our own datasets from donor liver tissue as a negative control. After filtering out probable germline mutations, we used two algorithms (PolyPhen and SIFT) to predict the effect of the missense mutations on protein function. We also determined whether these SNPs were likely to alter genes expressed in the hepatocytes by filtering out those genes not detected in liver tissue based on our previous data from microarray analysis.

When more DPMs (but not benign SNPs) normalised to the total number of non-synonymous SNPs, significantly were found in tumour tissue vs. non-tumour. This relationship was maintained when only liver-expressed genes were analysed, suggesting that damaging mutations affect cell phenotypes. Accumulation was observed despite the aetiology and known driver mutations in the tumour tissues, suggesting DPM accumulation is a general feature of carcinogenesis and could be used as a novel cancer biomarker.

POSTER 48

Wenhan Chen, Devika Ganesamoorthy and Lachlan Coin

DETECTING TUMOUR DNA IN BLOOD PLASMA USING ALLELIC IMBALANCE FROM READ COUNTDATA

ABSTRACT

Tumour cells undergo apoptosis release small amount of cell-free circulating tumour DNA (ctDNA) into blood throughout all stages of cancer. These ctDNA carry the somatic mutations which have accumulated in the tumour cells, including copy number alteration (sCNA) of whole chromosomes (aneuploidy) and chromosome arms. Aneuploidy is also a mark of high grade tumour development and poor survival prognosis. We hypothesised that next generation sequencing of cell free DNA in plasma may enable us to both detect the presence of ctDNA, and also characterise its copy number aberration profile. Many existing methods focus on using aberrant read-depth to infer the presence of copy number variation. We have investigated the use of aberrant allelic imbalance, as measured by the allele-specific read depths at SNP loci. We have adapted methods which combine population level phasing with read-backed phasing in order to maximise the information content in allelic imbalance. We demonstrate that allelic imbalance, modelled in this way, provides sufficient power to detect tumour DNA even when the mixture sample of tumour and normal DNA at a minimum 5% tumour cellularity. Moreover we are also able to pin down to the regions with copy number changes and predict possible aneuploidy.

POSTER 49

Erik Thompson, Melissa J Davis, Joseph Cursons, Karl Leuchowius, Mark Waltham, Sepideh Foroutan, Andrew Redfern, Eva Tomascovik-Crook, Bryce van Denderen, Tony Blick, Gayle Phillip, Edmund Crampin and Ian Street

THE ROLE OF MEK SIGNALING IN DRIVING MESENCHYMAL TRANSITION IN CELL LINE MODELS OF HUMAN BREAST CANCER METASTASIS

ABSTRACT

Epithelial mesenchymal transition (EMT) is the process whereby sessile, polarised epithelial cells alter the expression of key adhesion and regulatory molecules, and gain the ability to survive and migrate as single cells. While this process occurs normally in development we now recognize that metastasis has many elements in common with developmental EMT, which is subverted by carcinoma cells to allow metastatic spread. Cancer cells rarely undergo a full conversion to the mesenchymal phenotype, and instead appear to occupy a range of positions along a phenotypic spectrum between epithelial and mesenchymal states.

The EMPathy Breast Cancer Network (BCN) is a national collaborative effort including scientists, surgeons, medical oncologists and a consumer advocate investigating the role of EMP in breast cancer recurrence. The 7 thematic research projects of EMPathy BCN, including the 9 program-funded 'Satellite' projects, are aligned with the Cooperative Research Centre for Cancer Therapeutics (CTx) (www.cancercrc.com/index), so that any potential drug targets identified may progress into the CTx drug development program.

Multiple parallel approaches in the Target Discovery theme were used to identify candidate regulators and effectors of EMP. A total of 10 functional or gene expression experiments provided 7,950 significant events in any one system, which were cross referenced against 10 public breast cancer datasets relevant to EMP and/or breast cancer stem cells.

The Mitogen Activated Protein Kinase (MAPK) signaling cascade, which signals through MEK and ERK proteins, has previously been implicated in the regulation of EMT. In this work, we have used previously established breast cancer derived cell line models of epithelial-mesenchymal plasticity and applied RNA sequencing to measure transcriptional changes that occur as cells transition to a more mesenchymal phenotype.

We have identified co-ordinated changes in the expression of genes in many, although not all, MEK related signaling pathways, and identify modulation in the abundance of key miRs implicated in regulating genes across MEK related signaling pathways. Collectively, our results demonstrate that different MEK related signaling pathways are altered by the induction of EMT in our cell line models, depending on the position of the initial cell line on the epithelial-mesenchymal spectrum, and the type of stimulus used to trigger transition.

This work was supported in part by the EM*Pathy* National Collaborative Research Program funded by the National Breast Cancer Foundation, Australia.

POSTER 50

Dr Vincent Corbin, Dr Clare Fedele, Ms Samantha Boyle, Dr Mark Shackleton, Assoc. Prof. Anthony Papenfuss

DECIPHERING GENOMIC EVOLUTION IN HETEROGENEOUS TUMOURS

ABSTRACT

Cancer genomes can be highly unstable. Due to this instability, genomic evolution is typically rapid and highly heterogeneous within even a single tumour. The availability of genomic analysis methods now allows for the investigation of cancer genome evolution. Chromosomal SNP array and genome sequencing data provide information on dominant and non-dominant alleles, and can be used to study heterogeneity by analysis of B allele frequency, log R ratio and nucleotide mutation frequency.

We initially performed SNP array analysis on a primary malignant melanoma sample. We then derived single cells from the whole tumour and propagated these in mouse xenografts. Chromosomal SNP array analysis was then performed again on the tumour samples arising from the single cells.

Following comparison of the resulting copy number profiles, we were able to generate tumour-specific genograms, arising from the primary tumour through its daughter tumours and their subsequent heterogeneous genomic evolution. Comparison of copy number profiles was performed using a novel method based on the T-statistic between sliding windows on the log R ratio and B allele frequency channels.

Interrogation of the data arising from the tumour at multiple stages of evolution through the various daughter tumours illuminated specific evolutionary patterns, including the temporal pattern of genomic rearrangements, key genomic variations which occur independently within multiple daughter cell lines, and the pattern of change in the rate of tumour genome instability throughout multiple generations.

POSTER 51

Dr James Doecke, Dr Thierry Chekouo , Dr Kimberly Glass, Dr Marieke Kuijjer, Professor Kim-Anh Do

MIRNA-TARGET GENE RELATIONSHIP ANALYSES IDENTIFY KEY ONCOGENES FOR KIDNEY RENAL CLEAR CELL CARCINOMA PROGRESSION

ABSTRACT

Introduction: Although gene-cancer relationships are easily identifiable and usually strong, they are quite complex with multitudes of genomic elements responsible for the delineation of normal DNA replication, ultimately leading to cancer cell proliferation, and without treatment decreased survival time.

Methodology: Using The Cancer Genome Atlas (TCGA) data repository, we used level 3 Kidney Renal Clear Cell Carcinoma (KIRC) mRNA and miRNA expression data with the Passing Attributes between Networks for Data Assimilation (PANDA version 2) method[1]. We compared biological networks identified from two separate approaches: 1) network regulation via transcription factor (TF) – target gene associations, and 2) network regulation via miRNA - target gene associations. Furthermore, we use PANDA to combine both miRNA and TF's to assess the biological networks where target gene regulation is present from both sources. We then assessed these networks using the Ingenuity Pathway Analyses (IPA) tool.

Research design and hypotheses: Data was grouped from tissue samples where the clinical information indicated either FIGO stage 1/2, or FIGO stage 3/4 for comparison, such that both TF-target gene and miRNA-target gene networks were available for comparison between cancer stages. Our research hypothesis followed the premise that we would uncover differential networks for both TF-target gene and miRNA-target gene networks, between cancer stages. Moreover, we aimed to discover novel complex interactions between TF-miRNA and target gene related to cancer progression.

Data selection and pre-processing: From a total of 20,532 mRNA's, we selected only those genes with variance greater than the median (plus one standard deviation), that had less than 40% missing data, and that were significantly associated with survival time, leaving a total of 1,387 mRNA markers. mRNA data was log transformed, QQ-normalized and scaled prior to imputation using K-nearest neighbours algorithm. From the 805 available miRNA markers, we selected those miRNAs with predicted target genes (target gene selection using the approach in [2]) that matched the 1,387 mRNAs previously selected, leaving 458 miRNA markers to add into the PANDA method. Transcription factors (N=130) were selected using the list published in Glass et al., (2013).

Results: Taking two separate approaches to gene-regulator selection, we initially selected the top 50 genes (with their regulators) from each of the comparisons (six in total) for further network analyses with IPA. We then subtracted the PANDA edge coefficient from the FIGO stage 1/2 group from the FIGO stage 3/4 group, taking the top 50 maximally separated genes (and their regulators). Edge coefficients were greatest using the PANDA method with miRNA markers alone (without TF's), with a large proportion of strong relationships identified. Comparison of edge coefficients between FIGO stage groups using the miRNA data instead of TF's also identified the largest proportion of differential edges. Of the gene-regulator relationships with the strongest differences between cancer stage groups, we find 15 genes (CHRM1, CXCL5, CYP8B1, DSCAM, FAIM2, FUT3, GFAP, HAS2, IRF4, KCNMA1, MUC4, PEG10, PRDM16, PTPRZ1, SPIB) within a 30-gene network (Cell death and survival, cell cycle and cancer) with IPA. Genes at the centre of this network included TP53, CDKN1A, NFkB, CXCL8, MYC and ERK1/2. **Conclusions:** Extending the PANDA methodology to include miRNA-target gene regulation elucidated multiple genes in cancer and cell death/survival networks. Future work using this approach is aimed at identifying a pan-cancer pathway that may provide therapeutic targets for many different cancers.

POSTER 52

David Goode , Mandy Ballinger, Soroor Hediyeh Zadeh, Paul Leo , Matthew Brown, Ajay Puri, Isabelle Ray-Coquard, The International Sarcoma Kindred Consortium, David Thomas

AMPLICON-BASED TARGETED RESEQUENCING IDENTIFIES NOVEL GENETIC RISK FACTORS FOR SARCOMA AND AN ASSOCIATION BETWEEN GENETIC LOAD AND AGE OF ONSET IN SARCOMA PATIENTS

ABSTRACT

The risk of developing a sarcoma (a tumour of bone, muscle or connective tissue) has a strong heritable component, yet our knowledge of sarcoma risk genes remains incomplete. This study used the Haloplex amplicon sequencing assay to identify coding germline genetic variation in 100 targeted genes in 668 sarcoma patients recruited through the International Sarcoma Kindred Study (ISKS) who comprised a range of sarcoma subtypes and patient ages.

SNVs and short indels were filtered on the basis variant frequency, predicted functional consequence, sequencing quality metrics and annotations from the HGMDPro database. Our filters achieved >95% specificity for SNVs and enriched for deleterious germline coding variants, as evidenced by the association of the presence of selected variants with clinical relevant features, including early age of onset.

Comparison to an internal set of 227 controls sequenced using the same Haloplex capture reagents identified genes known to be associated with sarcoma risk as well as several novel candidate sarcoma risk genes. These findings were replicated two external exome sequencing data sets from populations of European heritage, using read depth to correct for coverage bias and synonymous variants to correct for ascertainment bias.

Furthermore, we uncovered evidence of positive epistatic interactions in germline genetic variation, as patients carrying multiple germline risk alleles had an increased risk of developing sarcoma earlier in life.

Our findings point to improvements in genetic screening for individuals at risk of early-onset cancers and demonstrate the utility of germline genetic variation in clinical decision-making.

POSTER 53

Ms Kaushala Jayawardana , Dr Sarah-Jane Schramm , Dr Varsha Tembe, Dr Samuel Müller, Professor Richard A. Scolyer, Professor Graham J. Mann, Associate Professor Yee Hwa Yang

MICRORNA PROGNOSTIC SIGNATURES IN METASTATIC MELANOMA

ABSTRACT

Predicting the survival outcome of patients is an important context in many biological studies as it aids to identify sub groups of patients that would ultimately lead to the development of individualized treatment therapies. With the recent advancements in high-throughput technologies and the generation of vast amounts of genetic information, the interest has been increasingly focused on using these data or a combination of these data to unveil intrinsic characteristics of patients to aid in disease prognosis. In this context, we analysed microRNA expression profiles of AJCC stage III melanoma patients. Outcome in surgically resected stage III disease is highly uncertain with a limited number of reliable and validated biomarkers, where the patient outcome is currently predicted using a limited set of clinical and histological features. We identified a 12-microRNA signature for the association of longer survival versus shorter survival. We assessed the prognostic utility of this signature in relation to top-ranking AJCC clinico-pathologic prognostic indicators and evaluated whether the signature adds predictive value when integrated with the same. Furthermore, we extended our study by conducting a formal and systematic cross-validation of the microRNA-based prognostic signatures in melanoma to date. Previous studies have systematically reviewed and/or cross-validated protein and mRNA biomarkers. However, to date there have been no such analyses applied to microRNA-based signatures of prognosis. Our findings provide direct insights into the translational potential of some signatures, while highlighting critical limitations of others.

POSTER 54

Anabel Kearney, Carrie Van Der Weyden & Silvia Ling

NEXT GENERATION SEQUENCING OF HAEMATOLOGICAL DISEASES IN A CLINICAL SETTING A PILOT STUDY

ABSTRACT

Molecular analysis of haematological diseases is of great benefit to clinicians and patients alike not only in the stages of initial diagnosis but also in the classification of risk groups and prognostic outcome. This pilot study examines the use of next generation amplicon sequencing as a method of mutation detection of five genes which have prognostic significance in acute myeloid leukaemia (AML) and three genes involved in the diagnosis of myeloproliferative disease (MPD).

The three-fold aim of this study was to: i) validate patient DNA panels, comparing the mutation status resulting from our institute molecular lab testing; ii) detecting new mutations in patients not screened before for particular MPD genes and in the form of base substitutions in the AML panel where the majority of mutations are detected and analysed by capillary electrophoresis, a method incapable of detecting such mutations and iii) challenge the present molecular testing scenario by detecting specific mutations for multiple diseases simultaneously using the same sequencing cartridge.

Methodology followed the Illumina 16S Metagenomic protocol for a 2 step PCR amplification method for amplicon and library preparation, substituting locus specific primers targeting the genomic areas of interest for those pertaining to the 16S ribosomal RNA gene. The targeted regions in this study

involved the FLT3, NPM1, CEBPA, WT1 and IDH1 genes in cytogenetically normal (CN) individuals with AML and the JAK2, CALR and MPL genes for patients suspected of MPD without probable reactive cause. Paired end Miseq target sequencing was performed and subsequent results analysed using Miseq Reporter.

Targeted next generation sequencing is highly applicable for mutation analysis in a clinical setting compared to standard laboratory procedures. It has the ability for high throughput and the convenience and capacity of screening individuals with different haematological diseases for pertinent, significant gene mutations in a single batch workflow.

POSTER 55

Farzaneh Kordbacheh, Nirav Bhatia, Camile S Farah

MOLECULAR PROFILE OF ORAL MUCOSAL LESIONS CHARACTERISED BY DIRECT TISSUE AUTOFLUORESCENCE

ABSTRACT

Loss of autofluorescence (LAF) using an optical imaging system (VELscope™) has been associated with potentially malignant disorders of the head and neck including oral epithelial dysplasia (OED) and squamous cell carcinoma (OSCC). This study utilised whole transcriptome analysis of oral mucosal biopsies to determine differential gene expression levels and their relation to LAF characteristics.

Whole transcriptome libraries were constructed from 43 frozen tissue biopsies with correlating LAF characteristics including 4 cases of OSCC, 10 cases of oral lichen planus (OLP), 15 cases of OED and 14 cases of benign oral epithelial hyperplasia (OEH). Barcoded libraries were then sequenced using an Ion Proton (Life Technologies™, CA, USA). All reads were mapped to the human reference genome, hg19, using Burrows-Wheeler Transform (BWA) algorithm. Sparse Partial Least Squares Discriminant Analysis (sPLS-DA) identified a potential gene list for each LAF pattern and correlated histology. Adjusted p-values were obtained from *limma*.

sPLS-DA revealed multiple up-regulated inflammatory and angiogenic related genes in LAF-BLANCHING (LB) lesions compared to NO-LAF (NL) and LAF-NO-BLANCHING (LNB) cases including *ICOS* in OEH and *TIE1* and *PECAM1* in OED lesions. Principal Component Analysis (PCA) clustered OLP samples closer to OSCC likely due to inflammatory markers. Overexpression of *CXCL13*, *FAM46C* and *ADAM28* were noted in LNB samples across all lesions (OEH□ OED□ OLP□ OSCC). This study reveals a relative molecular profile of characterised oral lesions that not only helps to identify the molecular pathways which may contribute to different phenotypes, but may also assist in stratifying lesions at risk of malignant transformation.

POSTER 56

Mr Marcus Lefebure, Dr Richard Tothill, Dr Jason Li, Dr Geoffrey Matthews, Dr Jake Shortt, Dr Edwin Hawkins, Dr Elizabeth Kruse, Dr Micah Gearhart, Professor Vivan Bardwell, Professor Ricky Johnstone

EXOME SEQUENCING REVEALS BCL6 CO-REPRESSOR (BCOR) AS A FREQUENTLY MUTATED TUMOR SUPPRESSOR GENE IN E-MYC LYMPHOMA

ABSTRACT

The Eμ-*Myc* mouse model employs a transgene mimicking the t(8;14) translocation found in Burkitt's lymphoma. Eμ-*Myc* mice develop Burkitt-like B-cell leukaemia/lymphoma with 100% penetrance. Progression from pre-malignant to clonal lymphoid neoplasia in the Eμ-*Myc* model ostensibly occurs when secondary somatic mutations disrupt the $p19^{ARF}$/p53 tumour suppressor axis.

We applied exome-sequencing to a cohort of 18 spontaneous murine Eμ-*Myc* lymphomas to determine novel somatic driver mutations capable of cooperating with *Myc*. In addition to exome-sequencing, we harvested blood

samples throughout the tumour progression at 2-week intervals to allow a temporal acquisition analysis of mutations that were detected at end-stage. Interestingly, concurrent driving mutations were evident in clonal tumours from the sequenced cohort, with *Cdkn2a* (encoding $p19^{ARF}$) deletions co-occurring with an activating *Kras* mutation. In addition to these mutations we identified novel protein-truncating mutations in Bcl6 co-repressor (*Bcor*), which are yet to be described in Eμ-*Myc* lymphoma. RNAi mediated knockdown of Bcor in Eμ-*Myc* foetal liver haematopoetic stem cells reconstituted into congenic recipient mice accelerated lymphomagenesis, validating *Bcor* as a tumour suppressor gene in this model.

This study challenges the current paradigm of tumour progression in Eμ-*Myc* as the data indicate that this system is perhaps a '3-hit model'. BCOR loss-of-function mutations have recently been identified in cytogenetically normal acute myeloid leukaemia (CN-AML), chronic lymphocytic leukaemia (CLL), retinoblastoma and medulloblastoma but it has not yet been validated as a genuine tumour suppressor gene. This study shows that *Bcor* loss-of-function mutations accelerate lymphomagenesis in a model of Burkitt-like leukaemia/B-ALL.

POSTER 57

Dr Evangelina Lopez De Maturana, Dr Antoni Picornell , Dr Manolis Kogevinas, Dr Mirari Marquez, Dr Alexandra Masson-Lecomte, Dr Montserrat Garcia-Closas, Dr Nathaniel Rothman Dr Francisco X Real , Dr Michael Goddard , Dr Nuria Malats

GENOME-WIDE MULTIMARKER MODEL TO PREDICT THE RISK OF BLADDER CANCER RECURRENCE

ABSTRACT

The aim of this study was to investigate if genome-wide common SNP profiles identified using Bayesian statistical learning improves the prediction of risk of recurrence in patients with non-muscle invasive bladder cancer when added to the usual clinico-pathological prognosticators (CPP).

We used data of 810 patients of the Spanish Bladder Cancer Study, who were followed up during 10 years and genotyped using the Illumina Infinium HumanHap1M array. The endpoint of interest was the time to first recurrence (TFR). Predictive ability of three models was evaluated: 1) including CPP, 2) including only SNPs and 3) combining CPP and SNPs.

We used Bayesian sequential threshold (BST) models to consider the time-to-event nature of the data, as well as censoring data. To accommodate and handle the large p small n problem, we combined BST with LASSO. Receiver operating characteristic curves with area under the curve (AUC) and determination coefficient estimates (R^2_{probit}) in a cross-validation scenario were used to evaluate the model predictions. Model with CPP classified better the patients at risk of having a TFR than the model with SNPs (0.62 vs. 0.55). In addition, it explained a larger proportion of the phenotypic variance than the model with SNPs (3% vs 1%). Although adding SNPs to the CPP improved the R^2_{probit} by 33%, it did not improve the classification (AUC) of new patients.

Our study shows the difficulty in predicting TFR using CPP and the little contribution of SNPs to predict TFR.

POSTER 58

Dr Mark McCabe, Dr Mark Cowley, Dr Sunita De Sousa ,Dr Melissa Vincent, Dr Reginald Lord, Dr Maija Corish, Dr Amber Johns, Professor Sandra O'Toole, Dr Ann McCormack, Associate Professor Marcel Dinger

STRATIFICATION OF THE CANCER GENOMIC LANDSCAPE THROUGH THE DEVELOPMENT OF A PAN-CANCER GENETIC SCREENING PANEL

ABSTRACT

Background and aims: Clinical use of cancer-screening panels is increasing for limited cancer types such as breast, lung and melanoma, due to their potential to identify causative/contributory genetic aberrations and the identification of new, or the repurposing of existing therapeutics, leading to better informed diagnoses with improved prognoses. The genetic basis underlying many rare, and/or poorly understood cancers, remains ill-defined. Patients afflicted with these conditions still face limited treatment options with poor prognoses. We have developed a pan-cancer genetic screening panel, which we aim to demonstrate the clinical utility for stratifying the genomic landscape of disparate cancer types, thereby improving treatment options and prognoses through discovering new or repurposing existing therapeutics.

Panel development: Our custom-designed gene panel through Roche/Nimblegen, combines 313 genes from various cancer screening panels including Foundation Medicine; enabling us to detect various germline and somatic mutations. This resulted in ~0.9Mb of target sequence and should provide excellent coverage across our 4372 targeted exons, after sequencing 24 samples in a single lane on an Illumina HiSeq 2500.

Methods: DNA was extracted from breast (phyllodes), pancreas, pituitary, lung and oesophagus tumour and/or whole blood samples and subjected to library preparation using the NEBNext Ultra kit. Sequencing data was aligned to the genome using bwa-mem, variants were identified using HaplotypeCaller, and prioritised using *in silico* methods, and ultimately, *in vitro* techniques.

Results: DNA from over 100 samples has been captured for sequencing analysis. The sequencing coverage was consistently high, with >90% of targeted regions sequenced to >100x. Our panel detected 700 variants, and genomic data is currently being analysed.

Conclusions: Our cancer panel is novel in its applicability to various cancer types and is designed to uncover new, causative genes as well as potentially identify alternative treatment options.

POSTER 59

Kelly Quek, Katia Nones, Ann-Marie Patch, Lynn Fink, Felicity Newell, David Miller, Karin Kassahn, John Pearson, Nic Waddell, Sean Grimmond

STRUCTURAL VARIATION IN PANCREATIC CANCER GENOMES

ABSTRACT

One of the hallmarks of cancer is the accumulation of genetic alterations which leads to uncontrolled growth of mutated cells. This alteration ranges from small mutations to large structural rearrangements. Structural variation includes duplications, deletions, insertions, translocation and other more complex genomics rearrangements. Despite a large number of studies and rapid progress of genomic technologies, the characterization of structural rearrangements and their underlying formation mechanisms are yet to be fully understood.

Here we present a comprehensive catalogue of somatic structural variation in 120 pancreatic primary tumours using high coverage whole genome sequencing data. Each cancer genome harbours a diverse array of somatic rearrangements and is different between individual genomes within cancer type. This suggests that different mechanisms may have present in the cancer genome or repaired these rearrangements.

Furthermore, we mapped the breakpoints to base pair resolution and have characterized the sequence context of the breakpoint junctions into five categories: blunt ends, microhomology, non-templated sequences, genomics shards and free DNA associated with mobile elements. Lastly, we described the patterns of joining broken chromosomal fragments in different subtypes of pancreatic cancer genomes to explore key biological insights to better understand the relationship between driving forces and their targets.

POSTER 60

Jason Ross, Ms Sue Mitchell , Dr Jason Ross , Dr Horace Drew , Mr Thu Ho , Mr Glenn Brown , Dr Graeme Young, Dr Trevor Lockett, Dr Susanne Pedersen, Dr Lawrence LaPointe, Dr Peter Molloy

COLORECTAL CANCER METHYLATION BIOMARKERS

ABSTRACT

The frequently methylated colorectal cancer (CRC) biomarker genes SEPT9, VIM1 and TMEFF2, have been developed into diagnostic assays. Despite their promise, there is considerable potential for the development of new DNA methylation biomarkers or panels to improve the sensitivity and specificity of current CRC detection tests.

To identify candidate genes methylated in a high fraction of CRCs we examined gene reactivation in cell lines following 5-aza/TSA treatment and also applied two novel in-house epigenome methods, SuBLiME and Bis-Tag, for methylation assessment in cell lines and clinical samples. We further evaluated differentially methylated genes using bisulfite amplicon deep-sequencing and quantitative methylation specific PCR.

Characterisaton of methylation in tumor, adenoma and non-neoplastic colorectal tissue and healthy donor peripheral blood led to discovery of a panel of 23 genes showing elevated DNA methylation in >50% of CRC tissue relative to non-neoplastic tissue. Six of these show very low methylation in non-neoplastic colorectal tissue and are candidate biomarkers for stool-based assays, while 11 genes have very low methylation in peripheral blood DNA and are candidate blood-based diagnostic markers. In particular, two methylated gene biomarkers, BCAT1 and IKZF1, looked most promising.

In a follow up study, this two-gene panel could detect 85 of 130 CRC cases in plasma specimens from 2139 volunteers scheduled for colonoscopy, which represents a sensitivity of 65%. In all non-neoplastic cases (n=1283), a specificity of 94.9% was observed.

Pilot release of the test commenced in July and is the first ever commercially available blood test for colorectal cancer in Australia.

POSTER 61

Dr Matthew Wakefield, Monique Topp, Valerie Heong, Alison Hadley, Australian Ovarian Cancer Study, Elizabeth Swisher, David Bowtell, Clare Scott

SELECTION RATHER THAN MUTATION DRIVES INCREASED PLATINUM RESISTANCE IN A PATIENT DERIVED XENOGRAFT MODEL OF OVARIAN CANCER

ABSTRACT

High Grade Serous ovarian cancer in clinical care is frequently treated with platinum based therapy.

Epithelial ovarian cancer response to platinum therapy is categorised as sensitive (a prolonged response), resistant (an initial response followed by an increase in tumour size), or refractory (increase in tumour size under therapy).

Ovarian cancer often becomes refractory under the selection of therapy, posing the question as to whether the refractory phenotype is selection of a subset of tumour cells or due to new acquired mutations.

To investigate the role of selection versus new mutation we have assessed the clonal diversity in high grade serous ovarian cancer patient-derived xenografts. These xenografts form part

of a consecutive cohort with characterised DNA repair gene mutations and platinum response status that is reflective of the clinical outcome of the patient (Topp et al 2014). Copy number variation was determined by whole genome sequencing to assess clonal diversity and its stability under serial xenografting and multiple rounds of therapy. We find that clonal diversity is maintained in our model system, and is stable in the absence of therapy. Under selection by platinum therapy some clones develop a refractory phenotype that is associated with a change in clonal diversity and prominence of a pre-existing chromosomal rearrangement, supporting clonal selection rather than contemporary mutation as the driver of the resistant to refractory phenotypic change.

The selected regions indicate several candidate genes which we plan to pursue with a CRISPR screen.

Topp et al. Mol Oncol. 2014 8:656-68.

POSTER 62

Dr Stephen Wong, Dr Andreas Behren, Dr Victoria Mar, Dr Katherine Woods, Dr Jason Li, Dr Claire Martin, Dr Karen Sheppard, Professor Rory Wolfe, Associate Professor John Kelly, Professor Jonathan Cebon

WHOLE EXOME SEQUENCING IDENTIFIES A RECURRENT RQCD1 P131L MUTATION IN CUTANEOUS MELANOMA

ABSTRACT

Melanoma is often caused by mutations due to exposure to ultraviolet radiation. This study reports a recurrent somatic C>T change causing a P131L mutation in the RQCD1 (Required for Cell Differentiation1 Homolog) gene identified through whole exome sequencing of 20 metastatic melanomas. Further screening in 715 additional primary melanomas revealed a prevalence of ~4%. This represents the first reported recurrent mutation in a member of the CCR4-NOT complex in cancer. Compared to tumors without the mutation, the P131L mutant positive tumors were associated with increased thickness (p= 0.02), head and neck (p=0.009) and upper limb (p=0.03) location, lentigo maligna melanoma subtype (p=0.02) and BRAF V600K (p=0.04) but not V600E or NRAS codon 61 mutations. There was no association with nodal disease (p=0.3). Mutually exclusive mutations of other members of the CCR4-NOT complex were found in ~20% of all subcutaneous melanomas from the TCGA dataset suggesting the complex may play an important role in melanoma biology. Mutant RQCD1 was predicted to bind strongly to HLA-A0201 and HLA-Cw3 MHC1 complexes and may therefore act as a neoantigen conferring a possible survival advantage to patients or be susceptible for immunotherapies in this patient subset.

POSTER 63

Dr Mark Cowley, Mr Kevin Ying, Dr Lisa Ewans, Ms Kerith-Rae Dias, Dr Warren Kaplan , Dr Timothy Furlong , Professor John Shine , Professor Michael Buckley , Dr Tony Roscioli, Associate Professor Marcel Dinger

IMPLEMENTING CLINICAL GENOMICS FOR DIAGNOSTICS AND RESEARCH

ABSTRACT

Next Generation Sequencing (NGS) has revolutionised genetic diagnostic testing, substantially improving diagnostic yields in patients with Mendelian disorders, from 7-12%, to on average 25-30%. NGS is also having a positive impact on cancer stratification, non-invasive prenatal testing, complex disease studies, and the emerging field of personal genomics. The intersection of medicine and genomics presents a unique opportunity for the diagnostic test to be used to drive research, both for improving diagnostic yields, and understanding the impact of genetic variation upon human health. Due to the substantial cost of developing and clinically validating a genomic test, we have taken the approach of focusing on whole exome and whole genome sequencing, with targeted analysis over known genes for diagnostic purposes, and then more exhaustive exome-wide analyses for gene and mutation discovery in undiagnosed cases.

In late 2012, we established the Kinghorn Centre for Clinical Genomics (KCCG) with the aim of implementing clinical genomics in Sydney. At the heart of the KCCG are two Illumina HiSeq 2500 sequencers that are used for rapid turnover exome sequencing, and more recently, one the world's first Illumina HiSeq X Ten sequencing systems, with capability of sequencing more than 300 whole human genomes per week. Much of our work over the past 18 months has been focused on building the infrastructure to support robust, reproducible, efficient clinical genomics pipelines, and working towards NATA accreditation. Clinically accrediting this process is challenging, in part due to the complexity of the analytical process, and the constantly evolving nature of the research-grade analytical components of the pipeline. To this end we have adopted rigorous best practice software engineering approaches, including Agile scrum methodologies that grant us the flexibility to continuously improve our processes and analytics.

To date, we have sequenced exomes from >200 patients, from a range of conditions, largely reflecting the undiagnosed caseload at the Sydney Children's Hospital, as well as a range of kidney disorders, osteoporosis, and dilated cardiomyopathy. From the Sydney Children's Hospital cohort, we performed whole exome sequencing on 122 individuals from 53 families (mostly intellectual disability and epileptic encephalopathy), and achieved a 38% diagnostic yield, with an additional 28% of cases with potential novel variants. Notably, early economic cost-benefit analyses are demonstrating that whole exome sequencing in the clinic is both cheaper and faster than traditional testing algorithms. For disorders where the degree of locus heterogeneity is small (ie <10 candidate genes), we have experienced even higher diagnostic yield, such as in cases of thin basement membrane disease, X-linked Alport's syndrome, and Nail-patella syndrome. We have struggled to diagnose patients with polycystic kidney disease using whole exome sequencing, due to pseudogenes confounding the alignment of short reads to the major candidate gene, *PKD1* (responsible for 85% of cases), which we are attempting to resolve with long range PCR.

I will present an overview of the systems that we are developing to enable clinical genomics at significant scale, as well as some early success stories and challenges from sequencing these clinical exomes and reflect on the possibilities presented by low-cost whole genome sequencing in the clinic.

POSTER 64

Dr Ruth MacKinnon, Dr Meaghan Wall, Associate Professor Lynda Campbell

GENOME CHARACTERISATION OF THE CELL LINE U937 PROVIDES FURTHER EVIDENCE FOR CENTROMERE CAPTURE AND DEMONSTRATES THE USE OF MULTIPLE GENOME TECHNOLOGIES TO EXAMINE EVOLUTION IN CELL LINES

ABSTRACT

We have carried out a detailed study of the genome of the myeloid cell line U937 using SNP (single nucleotide polymorphism) array with various FISH (fluorescence *in situ* hybridisation) techniques. Complex rearrangements were defined using SNP arrays to determine copy number aberrations and breakpoints, and different FISH techniques to determine the organisation of the aberrant segments within derivative chromosomes. While SNP array identifies copy number variation and breakpoints at high resolution, FISH can be used to characterise intact chromosomes and single cells, and localise highly repetitive DNA regions.

One benefit of SNP array is that the B allele frequencies allow the two chromosome homologues to be distinguished. For example, we showed that different chromosome 16 homologues were involved in an abnormal chromosome consisting of parts of chromosomes 4 and 16 and a chromosome consisting of 16p and 20p material. Similarly, the 1q material that had been translocated to chromosome 3 was not from the same homologue as the 1q material that had been lost from a del(1q).

This cell line provides an additional example of centromere capture, a recently described phenomenon in complex karyotypes. A short fragment containing the 11 centromere was joined to chromosome 16 and 20 material, enabling formation of a viable chromosome.

By considering the organisation of a complex abnormal genome using a variety of techniques, we have been able to make some inferences about how it has evolved. This strategy can help sort out evolutionary processes in cancer genomes with multiple rearrangement events.

POSTER 65

Gisela Mir, Tim Semple, Jason Li, Richard Tothill, Andrew Fellowes

LIBRARY PREPARATION, TARGETED-HYBRIDISATION CAPTURE AND SEQUENCING OF LIMITING AMOUNTS OF FFPE SAMPLES

ABSTRACT

Massively-parallel (next-generation) sequencing is now being rapidly adopted for the detection of pathogenic mutations in cancer patient samples. One of the great challenges faced in this process is the often limiting amounts and poor quality DNA that can be obtained from formalin-fixed and paraffin-embedded (FFPE) tissues. With the steady improvements to DNA sequencing and the chemistry used for library preparation, there are now a number of commercial options for processing and sequencing of FFPE samples. However, these methods still require thorough testing before they can be used routinely in a clinical setting under strict quality assurance guidelines. Targeted-hybridisation capture is routinely used in research and has been established as one of the most cost-effective methods to explore large genomic regions for diverse germ-line and somatic variations. We therefore aimed to explore the sensitivity, accuracy and reliability of different library preparation methods combined with hybridisation capture for sequencing of FFPE samples from pathology archives.

In this technical work we selected a good quality reference sample (NA12878, HapMap) and several tumor FFPE clinical samples to initially compare four library preparation methods (SureSelect XT, Ovation SP Ultralow, Kapa and ThruPLEX-FD) and two capture methods (SureSelect XT and NimbleGen SeqCap EZ) at different starting amounts of DNA (20 ng to 500 ng). We used a 3 Mb custom panel designed to capture 600 genes frequently mutated in cancer, some known fusions and one virus. Captured libraries were sequenced with an Illumina HiSeq2500 at 100 paired-end reads aiming a mean coverage >500x.

Although successful libraries were generated from 20 ng of DNA, using 50 ng or less DNA resulted in low complexity libraries and inefficient captures that produced high duplication levels and/or low on-target when aligned to the targeted regions. Underperformance at these conditions was also true for the good quality reference sample. When starting with 100 ng of DNA duplication rates were still higher than those expected in optimal situations but coverage was good for most of the targeted regions, with results very similar for the two capture platforms tested; at 300 ng we reached standard performance QC metrics. The highest variability in library efficiency between FFPE samples was irrespective of the library preparation method, emphasizing the importance of a good sample QC to estimate capture and sequencing success at an early stage.

Variant calling and copy-number variation analyses are being run to determine the sensitivity of the method. We aim to increase the number of samples and run validation tests to establish the combination of wet-lab and bioinformatics that best suits the use of targeted-capture and high-throughput sequencing for the study of FFPE samples in a clinical setting.

POSTER 66

Ms Mahtab Mirmomeni, Dr Kelly Wyres, Mr Benjamin Goudey, Dr Jeremy Wazny, Dr Andrew Mackinlay, Dr Priscilla Rogers, Dr Tom Conway

A CLINICAL READY SYSTEM FOR SCALABLE GENOMICS ANALYSIS

ABSTRACT

With the rapid drop of the cost of genome sequencing, its adoption as a routine procedure in clinical settings is fast approaching. It is essential to have platforms that can track and process the massive amount of data that will be produced. These systems require mechanisms for management of sequencing data, sample information (metadata), audit trail, analysis, visualization and higher-level analytics. In addition, they need to be easy to use for lab technicians. Given that genome sequencing might be used in multiple different domains, such as cancer genomics, pathogen genomics, etc, it is desirable to create systems that are flexible and not bound to a specific type of analysis. We introduce, SEGP, a Scalable Enterprise Genomics Platform, created to enable the routine use of genome sequencing in clinical settings. SEGP stores metadata as XML documents in a database. This enables higher-level analytics to be run on the data captured in the system e.g. for detection of anomalies or trends. All sequencing files are stored in a file system outside the database. Metadata is searchable. The result of running analyses on metadata and the associated sequences are captured as metaresult (human-readable and searchable) and result records (raw outputs). SEGP can easily adapted to different genomics applications. All data types for a particular use case, in addition to the workflows applicable to those data types, are configurable outside the platform using XML schema. This allows flexibility to adapt the platform to different settings without the need to modify source code.

POSTER 67

Mr John Pearson, Dr Ann-Marie Patch , Dr Katia None, Dr Kelly Loffler, Dr Henry Tang, Dr Andrew Barbour, Dr Nicola Waddell , Professor Sean Grimmond

NEXT-GENERATION SEQUENCING - BEYOND VARIANTS

ABSTRACT

Next-generation sequencing (NSG) has had a profound impact on genomics research and is making inroads in the clinic as part of the therapeutic decision making process. While there are numerous classes of genomic variants that can be called from NGS data - including nucleotide substitutions, insertions, deletions, copy number changes and structural variations – there are an increasing number of other classes of information that can be extracted from NGS data. These other classes of information can also play a major role in disease development and progression. The authors will discuss techniques and software tools (including a number developed by the authors) for looking at some of these including telomere quantification, pathogen screening, mitochondrial heteroplasmy and mutational signatures. We will highlight the importance of these approaches by applying them to cancer genomes from oesophageal and other cancer types.

POSTER 68

Mr Simon Sadedin, Dr Alicia Oshlack

CHALLENGES TO CNV DETECTION IN CLINICAL SETTINGS USING TARGETED HIGH THROUGHPUT SEQUENCING DATA

ABSTRACT

High Throughput Sequencing (HTS) is rapidly gaining clinical acceptance as a cost effective solution for diagnosis of highly penetrant but genetically heterogeneous diseases. Despite this success, the analysis methods applied are usually limited to detection of point mutations and small insertions and deletions. Detection of larger structural or copy number variants (CNVs) such as deletions, is often omitted despite such variants being equally deleterious and a common cause of many disorders. While many tools have been published in the literature for detecting structural variation in targeted sequencing, in practice, few laboratories are implementing these in their clinical sequencing analysis pipelines.

In this work, we apply a novel simulation method to explore the performance of CNV detection methods on real data. The simulation method inserts reads from real X chromosomes of male samples into female samples to simulate single copy deletions. We find that real world performance of published CNV detection methods appears to be highly sensitive to a range of factors including: method-specific tuning parameters, the exact sequencing technology in use, variability in the quality of data, and the number and type of samples compared in a batch. We conclude that in order for HTS CNV detection to become clinically accepted, methods must be developed that can work in a highly robust, self-calibrating fashion and that are well tuned to popular sequencing and capture platforms. Along these lines, we show results from our own method, Angel, which is specialized to achieve high accuracy and robustness on the HaloPlex targeted sequencing platform.

POSTER 69

Dr Jo-Ann Stanton, Dr Chris Rawle, Ms Christy Rand, Mr Chris Mason, Dr Richard Hall, Dr Jeremie Langlet, Dr Joanne Hewitt, Dr Hugo Strydom, Dr Angela Todd

A HANDHELD QPCR DEVICE FOR POINT-OF-CARE AND IN-FIELD DIAGNOSTICS

ABSTRACT

We have invented a Handheld Quantitative PCR (HHD qPCR) device that can operate on battery for up to six hours. The HHD qPCR device was tested using World Health Organization and IANZ accredited assays for *E.coli* STEC type strain NZRM3634, Influenza, Adenovirus, Enterovirus, Norovirus genogroup I & II, and Astrovirus. In side-by-side tests with larger laboratory based instruments and clinical samples the HHD qPCR was comparable to, and in one case better than, in-laboratory technology. Tests included measures for sensitivity, precision, and inter-assay variability.

The HHD qPCR detects both SYBR green and FAM reporter dyes. We have used DNA and RNA as reaction templates. Probe-based and intercalating dye assays have also been shown to perform successfully on the HHD qPCR device. The user interacts with the device via a tethered connection to a laptop computer. In addition, the HHD qPCR is wireless enabled to permit interaction with the operator via smart phone or tablet devices. Current reaction vessel volumes are equivalent to those found in a 96 well plate.

Diagnosis of infectious disease in the field or at the initial point-of-care brings the advantage of rapid infection containment, accurate diagnosis and immediate selection of the appropriate treatment. We see significant cost savings and benefits for the animal health sector with the technology quickly migrating to use in human health. For example, the HHD qPCR device could monitor for disease outbreaks as part of a national surveillance program. With the correct assay panel, our technology can determine the disease-causing species or reveal antibiotic resistance, all in real time either "cow-side" or in the remote clinic.

POSTER 70

Dr Siamak Tafavogh, Associate Professor Daniel Catchpoole, Associate Professor Paul Kennedy

GENERATING ACTIONABLE KNOWLEDGE FROM COMPLEX GENOMIC DATA FOR PERSONALISED CLINICAL DECISIONS

ABSTRACT

The human genome is complex. Within this complexity lies information about the individual. In conventional pathology clinicians group patients with similar cancer aggressiveness grades into broad clusters and deploy the same treatment regimens for all patients in a cluster. For example therapy of patients with low aggressiveness grade may be minor surgery or constant observation.

The main issue with conventional pathology is that deploying the same therapy for patients with the same aggressiveness grade might not lead to the same outcome because each individual has unique genotypic characteristics. Improved patient outcome requires the development of approaches that lead to personalised treatment strategies.

Computer Automated Diagnosis (CAD) links data mining and bioinformatics techniques to build a framework for personalized medicine. We present a CAD that indicates the genomic distances between patients, a concept we call 'similarity-space' which reflects the similarity or dissimilarity in patient genomic characteristics. It predicts the survival chance of new patients under the chosen therapy by computing the distance between the new and previously treated patients.

The CAD must work with imbalanced and high dimensional datasets. Thus, a multi-step algorithm is proposed and validated with preliminary results. Step1: we use resampling technique to balance the number of samples within different patients classes; Step2, we develop an approach using bootstrapping and random-forest to reduce the dimensionality; Step3, we develop a matrix-decomposition technique to map patients in the similarity-space. Our paradigm is childhood cancer and we aimed at testing the CAD in the context of acute lymphoblastic leukaemia.

POSTER 71

Ms Kim Ton, Macmil, S, Miller, A.L,Chua, E.W, Rand, C, Stanton, J.A, Kennedy, M.A

NEXT GENERATION SEQUENCING FOR PHARMACOGENETIC SCREENING

ABSTRACT

Pharmacogenetics (PGx) is contributing to the improvement of individualized medication and the prediction of clinical treatment outcome. With advances in this field, a wide range of PGx assays have been being developed and launched, ranging from low-throughput approaches that genotype single genetic mutations to high-throughput approaches to evaluate hundreds of genes simultaneously. The rapid evolution of next generation sequencing (NGS) makes it possible to sequence a large number of genes, up to the complete genome, in a single reaction. Hi-plex is an approach for generating NGS libraries via single multiplex PCR amplification (Nguyen-Dumont et al. 2014). This method uses automated primer design software and a simple, low-cost protocol for cost-effective and rapid performance. We are interested in exploring the application of next generation DNA sequencing methods to understand the impact of pharmacogenetic variation on drug responses and risk of adverse drug reaction. In this study, we use this the Hi-plex approach to develop an throughput assay for genotyping pharmacogenetic variants of *CYP2D6, CYP2C9, CYP2C19, DPYD, TMPT, UGT1A1, SLCO1B1, VKORC1*. These genes are involved in metabolism of a wide range of drugs, and some are also implicated in adverse drug reactions. We have focused on those genes with the best evidence of clinical utility (refer to PharmGKB VIP). Initially we are developing this assay for the Ion Torrent PGM platform. Our current iteration of this assay contains 15 amplicons for these eight genes, with indexing for 12 patients samples, and we aim to expand this further for more routine application.

Reference: Nguyen-Dumont, T., Pope, B. J., Hammet, F., Southey, M. C., & Park, D. J. (2013). A high-plex PCR approach for massively parallel sequencing. BioTechniques, 55(2), 69–74. doi:10.2144/000114052

POSTER 72

Dr Margaret Jordan, Dr Louise Laverick , Dr Melissa Gresle, Dr Dragana Stanley, Ms Letitia Smith, Dr Judith Field, Dr Laura Johnson, Dr Tim Spelman, Professor Helmut Butzkueven, Professor Alan Baxter

CAUSES OF MULTIPLE SCLEROSIS: A FUNCTIONAL GENOMICS APPROACH

ABSTRACT

Multiple Sclerosis (MS) is the most common disabling neurological disease affecting young adults in Western Society. Both environmental and genetic associations have been confirmed. To date, 110 strongly associated single nucleotide polymorphisms (SNPs) have been discovered but these represent common human genetic variations, and their individual association with MS is weak. The key challenge therefore is to identify the causal genes. One clue lies in the fact that many SNPs are in close proximity to known immune genes that could affect changes in immune cell function by changing their RNA expression and their encoded proteins. As it's highly likely that eQTLs and splice QTLs will be cell subset specific we wished to analyse expression levels in individual cell subtypes. While T cells likely play a critical role in MS pathogenesis, successful drug trials to date have targeted B-cells and monocytes in addition to T cells and there is also strong clinical and *in vitro* data identifying NK cell deficiencies in MS patients. We thus examined gene expression changes in 5 cell subtypes (CD4 and CD8 T cells, B cells, monocyes and NK cells) using Affymetrix Human Gene 1.0 ST arrays. Expression values were standardized across chips using RMA and quantile normalization and genes were ranked by expression difference significance by MWU and ANOVA. Expression differences of those genes adjacent to MS associated risk SNPs were considered. An investigation of candidate genes identified and the pathways to which they contribute is being undertaken to allow more targeted therapies to be developed.

POSTER 73

Nicholas Matigian, Elizabeth Mason, Dmitry Ovchinnikov, Othmar Korn, Ernst Wolvetang, Christine Wells

HOW DOES CHROMOSOME 21 TRISOMY ALTER THE GENE REGULATORY NETWORK OF NEURAL DIFFERENTIATION?

ABSTRACT

Down syndrome is a complex developmental and neurodegenerative disorder caused by trisomy of human chromosome 21 (HSA21). It occurs in approximately 1 in 700 live births, and is characterised by a hallmark facial features, growth delays, mental retardation and premature neural aging, heart defects as well as a range of other variable symptoms. These are assumed to arise from increased dosage of the HSA21 gene products, but this has not previously been modelled in the context of a developmental hierarchy of transcriptional regulation.

Using iPSc generated from two DS and two control patients we modelled early development of neurons. We stepwise differentiated these iPSc (Day0) into neuro-ectodermal cells (day6), neural stem cells (day12) and early neuronal progenitors (day 18) using an established neuronal differentiation protocols. At each stage we performed Capped Ananlysis of Gene Expression (CAGE) RNA-seq as part of the FANTOM5 project. This method produces gene expression profiles of all genes coupled with transcription start sites (TSS) position.

This work will identify key regulators in neuron differentiation that are sensitive to HSA21 trisomy altering the gene regulatory network of DS iPSc. The project uses a modular network approach to investigate the transcription factor network of *in vitro* iPSC->neural differentiation. Perturbation of the network imposed by Chr21 dosage will be assessed in terms of network membership, degree, or kinetic shifts, with the aim of predicting the molecular drivers of cellular defects that contribute to Down Syndrome pathologies. Analysis of this data has revealed the temporal expression pattern of HSA21 genes, with some genes escaping the dosage bias at some time points and not others. Additionally, non-HSA21 genes display dysregulation at different stages of development. We conclude that the effects of trisomy21 manifest in two forms (1) dosage of HSA21 elements that directly alter the phenotype of maturing neuronal cells, and (2) temporal modulation of HSA21 elements that directly or indirectly alter the normal developmental trajectory.

POSTER 74

Associate Professor Leanne Dibbens, Michael G Ricos, Sarah E Heron, Bree L Hodgson, Ingrid E Scheffer, Samuel F Berkovic

EPILEPSY: NEW GENES AND OLD PATHWAYS IN A COMPLEX DISORDER

ABSTRACT

Epilepsy is the most common serious neurological disorder affecting 2% of the world population. More than 30 different epilepsy syndromes have been defined based on various clinical criteria. Around 25% of cases are symptomatic and are associated with due to brain insults eg. stroke, tumour or brain injury but 70% of all epilepsies have a genetic cause making them ideal for molecular genetic analyses. Over the past 15 years our group has collected more than 10,000 epilepsy cases from around the world with linked clinical information into a bio-bank. We have been analyzing this cohort with evolving genomic technologies and have identified a number of genes as the cause of "monogenic" epilepsies. Where available we have used linkage analysis and next generation sequencing approaches to identify candidate genes in epilepsy and its co-morbidities including intellectual disability, autism spectrum disorders and psychiatric features.

A number of ion channels genes including nicotinic acetylcholine receptors (e.g. CHRNA4), and sodium gated potassium channels (e.g. KCNT1) have formed the

majority of "epilepsy genes" discovered. We have identified genes involved in secretory pathways (e.g. PRRT2, GOSR2, SCARB2). Interestingly we have also Identified structural proteins (e.g. PCDH19). Despite the larger number of 'epilepsy' genes identified these only account for a small proportion of the cases of epilepsy studied. Significantly, we have now identified signal transduction molecules from the mTOR pathway in focal epilepsy, finding DEPDC5 mutations in 12% of cases of familial focal epilepsy studied, making it the most common known cause of epilepsy.

POSTER 75

Dr Rima Chaudhuri, Dr Poh Khoo, Dr Jean Yang, Dr Jagath Junutula, Dr Ganesh Kolumam, Dr Zora Modrusan, Dr Kyle Hoehn, Dr Samantha Hocking, Dr Jerry Greenfield, Professor David James

CROSS-SPECIES SYSTEMS ANALYSIS OF A METABOLIC DISEASE TO FIND A GENE SIGNATURE PREDICTIVE OF INSULIN RESISTANCE

ABSTRACT

Insulin resistance (IR) is perhaps the best predictor of future development of type 2 diabetes (T2D). Identification of molecular signatures that can identify individuals with higher risk of developing T2D could enable early stage intervention. Gene expression data could provide an ideal tool for identification of such a molecular signature, but human gene expression (GE) data is inherently noisy and highly variable. To overcome the lack of power, we integrated human data with more stable data from a model organism and thereby developed a cross-species (human-mouse) analysis platform. We applied this new approach to identify a gene signature that could differentiate insulin resistant from insulin sensitive individuals with improved prediction accuracy (~20% increase) compared to standard clinical measures (OCM). This signature also identified beta-catenin and JAK1 as novel 'connection hubs' to the Insulin Signaling Pathway whose GE pattern highly correlated to extensive metabolic phenotypic measures of insulin sensitivity in humans. Inhibiting these proteins impaired insulin-stimulated glucose uptake in L6 myotubes, confirming their role in insulin action. These data indicate the potential utility of using systems biology approaches to segregate individuals early based on differential diabetes risk.

POSTER 76

Nicole Cloonan, Hilary C Martin, Shivangi Wani, Anita L Steptoe, Keerthana Krishnan, Katia Nones, Ehsan Nourbakhsh, Alexander Vlassov, Sean M Grimmond

IMPERFECT CENTERED MIRNA BINDING SITES ARE COMMON AND FUNCTIONAL

ABSTRACT

MicroRNAs (miRNAs) bind to mRNAs and target them for translational inhibition or transcriptional degradation. It is thought that most miRNA-mRNA interactions involve the seed region at the 5□ end of the miRNA. The importance of seed sites is supported by experimental evidence, although there is growing interest in interactions mediated by the central region of the miRNA, termed centered sites. To investigate the prevalence of these interactions, we apply a biotin pull-down method to determine the direct targets of ten human miRNAs, including four isomiRs that share centered sites, but not seeds, with their canonical partner miRNAs.

We confirm that miRNAs and their isomiRs can interact with hundreds of mRNAs, and that imperfect centered sites are common mediators of miRNA-mRNA interactions. We experimentally demonstrate that these sites can repress mRNA activity, typically through translational repression, and are enriched in regions of the transcriptome bound by AGO. Finally, we show that the identification of imperfect centered sites is unlikely to be an artifact of our protocol caused by the biotinylation of the miRNA. However, the fact that there was a slight bias against seed sites in

our protocol may have inflated the apparent prevalence of centered site-mediated interactions.

Our results suggest that centered site-mediated interactions are much more frequent than previously thought. This may explain the evolutionary conservation of the central region of miRNAs, and has significant implications for decoding miRNA-regulated genetic networks, and for predicting the functional effect of variants that do not alter protein sequence.

POSTER 77

Dr Thomas Conway, Dr Kelly Wyres, Dr Keira Cohen, Dr Tal El-Hay, Dr Omer Weissbrod, Dr William Bishai, Dr Alex Pym

DERIVING MYCOBACTERIUM TUBERCULOSIS SPOLIGOTYPES FROM NGS DATA

ABSTRACT

Spoligotyping is a genotyping assay applicable for *Mycobacterium tuberculosis* (*M. tb*), the causative agent of tuberculosis. It is used for identifying the relationships between isolates, tracking the spread and evolution of *M. tb*. Traditional spoligotyping is a laboratory technique, whereby genomic DNA is assayed for hybridisation to a set of 43 nucleotide probes, each 25 base pairs in length. Researchers are increasingly moving away from traditional genotyping techniques in favour of high resolution next-generation sequencing (NGS), which is also generating much interest in the clinical and public health microbiology space. However, derivation of traditional genotyping information from these data remains key for backwards compatibility and contextualisation of results.

We present a method for deriving spoligotypes from NGS data. We use an alignment-free method that avoids the cost and biases associated with mapping or *de novo* assembly, and results in an assay that closely mirrors the original hybridization based assay. Isolates for which experimental spoligotypes and NGS data are both present were not available, so we demonstrate the efficacy of the method on a collection of 86 *M. tb* clinical isolates from South Africa, sequenced as part of another study. We show that the lineages that the spoligotypes associate to are consistent with a phylogenetic tree derived from variation across the whole genome. Furthermore, spoligotype-defined lineage assignments were consistent with those known for publicly available genome sequences included in the phylogenetic analysis. Our method is efficient and robust, meaning it is appropriate for research deployment and public health use.

POSTER 78

Harriet Dashnow, Susan Tan, Dr Debjani Das, Professor Simon Easteal, Dr Alicia Oshlack

GENOTYPING MICROSATELLITES IN NEXT-GENERATION SEQUENCING DATA

ABSTRACT

Microsatellites are short (2-6bp) DNA sequences repeated in tandem, which make up approximately 3% of the human genome. These loci are prone to frequent mutations and high polymorphism. Dozens of neurological and developmental disorders have been attributed to microsatellite expansions. Microsatellites have also been implicated in a range of functions such as DNA replication and repair, chromatin organisation and regulation of gene expression.

Traditionally, microsatellite variation has been measured using capillary gel electrophoresis. In addition to being time-consuming, and expensive, this method fails to reveal the full complexity at these loci because it cannot detect SNP polymorphisms and compound microsatellites.

Next-generation sequencing has the potential to address these problems. However, determining microsatellite lengths using next-generation sequencing data is difficult. In particular, polymerase slippage during PCR amplification introduces stutter noise. A small number of software tools claim to genotype simple microsatellites in next-generation sequencing data, however they fail to address

the issues of SNPs and compound repeats, and they tend to provide only approximate genotypes.

We have developed a microsatellite genotyping algorithm that addresses these issues, providing high accuracy as well as more detailed analysis of microsatellite loci. We have validated it using high depth amplicon sequencing data of microsatellites near the *AVPR1A* gene. We found high concordance between our algorithm and repeat lengths obtained by electrophoresis, manual inspection and Mendelian inheritance. By subsampling the reads, we found that our model is accurate to within one repeat unit down to coverages that we would expect in standard exome sequencing.

POSTER 79

Dr Saskia Freytag, Dr Juliane Manitz, Professor Dr Martin Schlather, Professor Dr Thomas Kneib, Professor Dr Christopher Amo, Professor Dr Angela Risch Dr Jenny Chang-Claude, Professor Dr Joachim Heinrich, Professor Dr Heike BickebÃ¶ller

IDENTIFYING DISEASE-ASSOCIATED PATHWAYS IN GENOME-WIDE ASSOCIATION STUDIES VIA A NETWORK-BASED KERNEL APPROACH

ABSTRACT

Biological pathways provide rich information and biological context on the genetic causes of complex diseases. The logistic kernel machine test integrates prior knowledge on pathways in order to analyze data from genome-wide association studies (GWAS). In this study, the kernel converts the genomic information of two individuals into a quantitative value reflecting their genetic similarity. With the selection of the kernel, one implicitly chooses a genetic effect model. Like many other pathway methods, none of the standard kernels accounts for the topological structure of the pathway or gene-gene interaction types. However, there is evidence that connectivity and neighborhood of genes are crucial in the context of GWAS, because genes associated with a disease often interact. Thus, we propose a novel kernel that incorporates the topology of pathways and information on interactions. Using simulation studies, we demonstrate that the proposed method maintains the type I error correctly and can be more effective in the identification of pathways associated with a disease than non-network-based methods. We apply our approach to genome-wide association case-control data on lung cancer and rheumatoid arthritis. We identify some promising new pathways associated with these diseases, which may improve our current understanding of the genetic mechanisms.

POSTER 80

Miriam Fanjul Fernandez, Peter Hickey, Vesna Lukic, Natasha Brown, Greta Gillies, Sarah Wilson, Martin Delatycki, Ingrid Scheffer, Melanie Bahlo, Paul Lockhart

GENE DISCOVERY IN LARGE MULTIPLEX FAMILIES WITH AUTISM SPECTRUM DISORDERS

ABSTRACT

Autism spectrum disorders (ASD) are common neuro-developmental disorders, defined by impairment in language and social interaction, that affect ~1:100 individuals. ASD typically display complex inheritance with a multifactorial basis, but despite the significant heritability, the genetic basis underpinning the disease remains largely unexplained. Moreover, up to 20% of relatives of individuals with ASD display the "broader autism phenotype" (BAP), characterised by less severe deficits in one or more of the three core ASD domains.

Genetic studies have demonstrated that next-generation sequencing (NGS) technology can be a powerful tool to identify the causative genes. However, the high genetic heterogeneity characterising ASD along with studies focusing on small multiplex kindreds may explain previous unsuccessful molecular studies. We have employed a novel approach that analyses the BAP as marker of an ASD genetic variant in large multiplex families that

are more likely to be genetically homogenous and suitable for classical linkage approaches to gene discovery. To date, more than 40 individuals from two large pedigrees have been exome sequenced and genotyped using Illumina platforms. Reads were aligned to the reference genome (GRCh37) using Novoalign and single point mutations and indels were called simultaneously across all sample BAMs with GATK-Haplotype Caller and annotated with Annovar. On average, individuals have 300,000 raw variants, including artefacts and non-relevant variants. After applying numerous filtering steps based on allele frequency, biological consequences and inheritance models, we have selected several candidate genes for further evaluation. Finally, analysis of the segregation of the mutation within each family and functional studies will allow us to identify the causative ASD variant.

POSTER 81

Nicole Lake, Sze Chern Lim, Katherine R. Smith, David A. Stroud, Alison G. Compton , Luke C. Gandolfo , John Christodoulou, Michael T. Ryan, Melanie Bahlo, David R. Thorburn

THE GENETIC BASIS OF LEIGH SYNDROME, THE MOST COMMON MITOCHONDRIAL DISORDER AFFECTING CHILDREN: A COHORT STUDY

ABSTRACT

Leigh (-like) syndrome (LS) is the most common paediatric clinical presentation of inherited mitochondrial energy generation disorders. This progressive neurodegenerative disease is characterised by the development of bilateral symmetrical lesions in the brainstem and basal ganglia. However LS is clinically and genetically heterogenous, and can be caused by mutations in over 50 different genes. We present a cohort of 113 patients (95 pedigrees) with a clinical, radiological and genetic diagnosis of LS. Mutations were identified through diagnostic screening or targeted massively parallel sequencing (MPS) research studies. Presentation age was available for >80% of patients, with a median age of onset of 8 months. Pathogenic mutations in seventeen different disease genes and a mitochondrial DNA deletion were identified. These mutations were inherited in maternal (43%), autosomal recessive (45%) or X-linked (12%) manners. The identified disease genes encoded subunits or assembly factors of oxidative phosphorylation (OXPHOS) complexes I (35%), IV (23%), V (20%) or pyruvate dehydrogenase (12%). Molecular defects in genes associated with a combined OXPHOS deficiency were also observed in 10% of patients. We recently characterised PET100 as a novel LS disease gene encoding a complex IV biogenesis factor. A pathogenic homozygous c.3G>C mutation predicted to abolish the initiation codon was identified in ten patients (8 families) of Lebanese descent. Lentiviral-mediated overexpression in patient fibroblasts restored complex IV assembly and activity, confirming it as a bona fide disease gene. Our results highlight the significant genetic heterogeneity underlying LS. Future MPS based studies are anticipated to reveal additional novel LS disease genes.

POSTER 82

Victoria M. Perreau, Alberto Capurro, Liviu Bodea, Patrick Schaefer, Ruth Luthi-Carter

COMPUTATIONAL DECONVOLUTION OF GENOME WIDE EXPRESSION DATA FROM PARKINSON'S AND HUNTINGTON'S DISEASE TISSUES USING POPULATION-SPECIFIC EXPRESSION ANALYSIS.

ABSTRACT

The characterization of molecular changes in diseased tissues gives insight into pathophysiological mechanisms and is important for the development of treatments. Genome-wide gene expression analysis has proven valuable for identifying biological processes in neurodegenerative diseases using post mortem human brain tissue and numerous datasets are publically available. However, many studies utilize heterogeneous tissue samples consisting of multiple cell types, all of which contribute to global gene expression values, confounding biological interpretation of the data. Elucidation of the

cell-type-specific contributions to expression profiles enables better understanding of the underlying processes. For example, changes in numbers of neuronal and glial cells occurring in neurodegeneration confound transcriptomic analyses, particularly in human brain tissues where sample availability and control is limited.

To this aim, we have applied our recently published computational deconvolution method, population specific expression analysis (PSEA). PSEA estimates cell-type-specific expression values using reference expression measures, which in the case of brain tissue comprises mRNAs with cell-type-specific expression in neurons, astrocytes, olygodendrocytes and microglia. As an exercise in PSEA implementation and hypotheses development regarding neurodegenerative diseases, we applied PSEA to Parkinson's and Huntington's disease (PD, HD) datasets. Genes identified as differentially expressed in specific cell types by PSEA were validated using external data. Network analysis of differentially expressed genes and Annotation Clustering (DAVID) elucidated molecular processes implicated by differential gene expression in specific cell types. The results of these analyses provided new insights into the implementation of PSEA in brain tissues and additional refinement of molecular signatures in human HD and PD.

POSTER 83

Dr Michael Ricos , Dr Sarah Heron, Dr Ingrid Scheffer, Ms Bree Hodgson, Dr Simone Mandelstam , Dr Douglas Crompton, Dr Francesca Bisulli, Dr Paolo Tuniper, Professor Samuel Berkovic, Associate Professor Leanne Dibbens

MUTATIONS IN THE MTORC1 REGULATOR DEPDC5 ARE A MAJOR CAUSE OF LESIONAL AND NON-LESIONAL FOCAL EPILEPSY

ABSTRACT

We set out to find the genetic cause of focal epilepsy in a family with individuals that had lesional and non-lesional focal epilepsy, as detected by MRI analysis. The focal epilepsy showed autosomal dominant inheritance and linkage analysis failed to identify any linkage region.

We carried out exome sequencing on two individuals from the family who were affected with focal epilepsy and displayed abnormal MRI. The sequence data was analysed using an in-house bioinformatic pipeline. Candidate causative genetic variants were identified and validated by Sanger sequencing. These were analysed for co-segregation with affected status and assessed for likely pathogenicity.

We identified a mutation in DEPDC5 as being causative of both lesional and non-lesional focal epilepsy in the family. We also identified two other families with DEPDC5 mutations who also had mutation-positive individuals with lesional and non-lesional focal epilepsy. DEPDC5-associated malformations included bottom-of-the-sulcus dysplasia (3 members from 2 families) and focal subcortical band heterotopia (1 individual). We show that mutations in DEPDC5 cause familial cases of focal epilepsy associated with structural lesions. Previously we found that mutations in DEPDC5 caused familial cases of non-lesional focal epilepsy. We therefore now show that lesional and non-lesional epilepsy can have a shared genetic aetiology. This challenges previous dogma of lesional and non-lesional epilepsy being regarded as distinct entities. DEPDC5 is part of the GATOR1 complex, which negatively regulates mTORC1 signalling and DEPDC5 mutations are associated with brain malformations similar to Tuberous Sclerosis patients where TSCI and TSCII are mutated.

POSTER 84

Lyndal Henden, Melanie Bahlo, Marie Shaw, Alison Gardner, Jozef Gecz

INFERRING IDENTITY BY DESCENT ON THE X CHROMOSOME

ABSTRACT

Identity by descent (IBD) detection has played a crucial role within the field of human disease mapping. The idea behind IBD in disease mapping is that if a disease is heritable, then affected individuals in the same family are likely to have the same disease-causing variant. Affected individuals will then also have a shared region of IBD around the causal variant that will not be present in unaffected individuals. Substantial progress has been made with regards to IBD methodology such as the inclusion of genotyping errors and accounting for linkage-disequilibrium, driven by the generation of denser marker data such as that generated by next generation sequencing (NGS), however little to no work has been done on the X-chromosome. Our work focuses on the "forgotten X" and as such we have developed a Hidden Markov Model for inferring IBD sharing on the X-chromosome. This method is designed for SNP array genotype data, however it can also be applied to NGS data once the data has been formatted as SNP array data. Using exome sequencing data from individuals with X-linked intellectual disability, we show that our method can narrow the search space for causal variants on the X-chromosome as well as consolidate evidence that a variant is causal.

POSTER 85

Ms Tracey Van Stijn, Ms Shannon Clarke, Mr Rudiger Brauning, Mr Ken Dodds, Mr John McEwan

GENOTYPING-BY-SEQUENCING (GBS): COMPARISON TO SNP CHIPS AND WHOLE GENOME SEQUENCING

ABSTRACT

Genotyping-by-Sequencing (GBS) has the potential to be a cost effective, reproducible and high-throughput SNP genotyping method. Although we have been investigating GBS in a number of livestock and aquaculture species, we have concentrated largely on sheep. To establish the accuracy and reproducibility of GBS, and to allow forward genotyping capability, SNPs located in GBS targeted genome regions were included on the sheep High Density (HD) Illumina SNP chip. We have compared genotype calls between the two genotyping methods as well as whole genome sequencing (between 10x – 20x genome coverage).

POSTER 86

Ms Anita Goldinger, Dr Jian Yang, Dr Anjali Henders, Dr Allan McRae, Professor Grant Montgomery, Professor Peter Visscher, Professor Greg Gibson, Dr Joseph Powell

EXTENSIVE GENETIC CO-REGULATION DRIVES HIGHLY CORRELATED MODULAR STRUCTURES IN TRANSCRIPTOMIC DATA

ABSTRACT

Investigating the genetic co-regulation of RNA transcripts can elucidate the complex genetic architecture governing the translation of gene products into higher-order phenotypes. In a previous study, we identified nine modules of correlated genes, which were representative of distinct changes in immunological function due to infection or geographical location. Here, we report results from our examination into the genetic co-regulation of these modules. The phenotypic correlations between module probes (ranging from 0.35 to 0.89) are significantly higher than estimates obtained from randomly sampled probes. The probe heritability (h^2), estimated using GREML analysis, was high with > 90% of the probes demonstrating a significant h^2 between 0.19 and 0.65. To calculate the amount of covariance attributed to common genomic control, we calculated the genetic correlation (r_G) using bi-variate GREML, for all pair-wise combinations of module probes. The average r_G estimate of these probes was higher than the phenotypic correlations, ranging from 0.56 to 0.96. The high r_G suggests that common loci regulating module probes are driving the high phenotypic correlations. A significant number of trans-eQTLs were shared between probes in each module. These eQTLs were located in genes involved in hematopoiesis and were also enriched for previous GWAS hits with various blood cell related disease. Our results demonstrate that these modules are under tight genetic regulation, which facilitates their coordinated response to environmental pressures. This information has important application in the functional annotation of genomic loci, building causal networks driving disease and exploring the basis of disease susceptibility between individuals.

POSTER 87

Ms Beth Signal, Dr Brian Gloss, Mr Dominik Kaczorowski, Mr Seth Cheetham, Ms Franziska Gruhl, Dr Marcel Dinger

THE DISCOVERY OF NOVEL LINCRNA WITH POTENTIAL ROLES IN MESC DEVELOPMENT

ABSTRACT

A large proportion of the genome can be transcribed as long non-coding RNA (lncRNA), a class of non-coding RNA with demonstrated biological importance. Despite multiple available computational tools being available, the categorisation of these transcripts as non-coding from sequence information alone remains difficult.

It is also highly likely that novel lncRNAs remain to be found due to their characteristic expression pattern of low levels combined with high cell-type specificity and rapid turnover. We hypothesised that employing higher temporal resolution RNA-Seq time courses would allow us to capture novel, transiently expressed or highly specific, lncRNA transcripts.

We thus sought to identify potentially novel intergenic lncRNAs (lincRNAs) important in cellular differentiation using RNA-Seq data from a six-hourly time course over six days of mouse embryonic stem cell (mESC) development. Genome guided transcriptome reconstruction was employed to generate a 'novel transcriptome' of mESC differentiation. Transcripts from this were then subjected to a lincRNA detection pipeline established by comparing the performance of multiple computational tools for classifying transcripts as coding or non-coding on annotated mouse transcripts. This method facilitated the discovery of hundreds of novel lincRNA with potential roles in mESC development. Insights into potential functions of these novel transcripts were gained by comparing expression profiles to those of known transcripts and detecting expression changes in other developmental transcriptome datasets.

POSTER 88

Suzanne Butcher, Edward Huang, Ashley Waardenberg, Othmar Korn, Rowland Mosbergen, Tyrone Chen, Kelly Hitchens, Dipti Vijayan, Anthony Beckhouse, Mark Walker, Tim Barnett, James Frazer, Antje Blumenthal, Matt Sweet, The FANTOM5 consortium and Christine Wells

DIVERSIFICATION OF INNATE IMMUNE RESPONSES BY TRANSCRIPTIONAL MECHANISMS

ABSTRACT

It has long been established that innate immune cells elicit a broad inflammatory program as a rapid first-line of defence against pathogens. Less well understood are pathogen-specific responses, which vary by the type of inflammatory molecule, amplitude and/or kinetics of inflammatory response. The ability of host innate immune cells to direct specialist inflammatory responses has now been described for numerous pathogens - across many taxonomic levels, right down to individual species and strains of pathogen. As the spectrum of innate cell phenotypes produced by these inflammatory programs drives long term adaptive immunity, it is important that the factors underlying their diversity are understood. Current pathway or network models of innate immune signalling do not reflect phenotypic diversity, or provide a mechanistic explanation for the evolution of divergent molecular responses from convergent signalling pathways. While this convergence is driven from evolutionary conservation of the genome, transcriptomic diversification mechanisms may provide a means to contribute to diversity in innate immune responses.

We have used Cap Analysis Gene Expression (CAGE) to characterise common and unique components of monocyte transcriptional programs elicited by representative pathogens. Within these programs we have identified candidate genes subject to regulation by multiple transcription start sites (TSSs). Our findings suggest that alternate TSSs provide scope to diversify innate immune responses by producing variant protein products, and impacting post-transcriptional transcript fate. These findings indicate TSS engagement patterns are an exciting regulatory layer which should be considered when deciphering phenotypic diversity in innate immune responses.

POSTER 89

Dr Brian Gloss, Ms Bethany Signal, Mr Dominik Kaczorowski, Mr Seth Cheetham, Ms Franziska Gruhl, Associate Professor Marcel Dinger

COMPLEX CODING AND NONCODING TRANSCRIPTIONAL DYNAMICS IN EARLY STEM CELL DEVELOPMENT

ABSTRACT

Low cost-high throughput sequencing technology is allowing researchers to revisit questions previously too costly or complex to answer. One example of this is transcriptional profiling studies of development, which are typically sampled at only a small number of time points. In the case of complex mammalian embryonic stem (ES) cell development, such profiling is likely misses key developmental transitions. Furthermore, as RNA signals and especially noncoding RNAs can be extremely labile, it is probable that our understanding of transcriptional events underlying normal development –and consequently developmental disease- is missing significant detail. We have performed RNA sequencing in unprecedented temporal detail on the first 5 days of mouse ES cell development and are unraveling exquisite complexity of coding and noncoding gene dynamics.

This temporal detail has allowed us to observe previously unknown variation in expression of well-characterised developmental genes as well as significantly increased regulatory complexity. We have identified hundreds of rapidly cycling (<24hr cycle) coding and non-coding transcripts in ES development. Additionally. We have also identified transcriptional networks of protein coding

genes (enriched for cell cycle, focal adhesion and transcription factors) responding to expression changes of novel noncoding RNAs. Finally, this data is enabling unprecedented capacity to unravel driver/carrier transcriptional dynamics at complex gene loci (bi-directional/antisense).

Fine temporal resolution has highlighted important gene expression changes in early differentiation and identified many noncoding transcripts with a potential driver role in cellular development. It is becoming possible to elucidate the complexities of transcriptional control during cellular differentiation.

POSTER 90

Dr Yunshun Chen, Professor Gordon Smyth

DETECTING DIFFERENTIAL EXON USAGE IN RNA-SEQ

ABSTRACT

Alternative splicing is a process where exons are differentially combined or skipped, resulting in multiple protein isoforms encoded by a single gene. It generates diverse transcripts and provides an opportunity for gene regulation. The proteins translated from alternatively spliced mRNAs will contain different or even opposite biological functions. It has been discovered that more than 95% of human multi-exon genes express multiple splice isoforms. Alternative splicing can affect various functions in cellular processes, tissue specificity, developmental states and disease conditions. Recent high throughput technologies, such as RNA-Seq, have been proven to be very powerful tools in studying alternative splicing. Meanwhile, it is also very tempting and challenging for statisticians and bioinformaticians to develop statistical methods for detecting alternatively splicing events. Here we present a method for testing for differential exon usage between different experimental conditions or different genetic backgrounds. In particular, we compare the change of the expression level of each exon to the change of the expression level of the gene containing that exon under a certain comparison. Similarly to the gene level analysis of RNA-Seq data, we use negative binomial distribution to model exon counts. An empirical Bayes information sharing strategy is applied to estimate the variation between biological replicates. Statistical tests are performed at both the exon level and the gene level. The method has been implemented in the edgeR package.

POSTER 91

Dr Wei Shi, Dr Bhupinder Pal, Prof. Geoffrey Lindeman, Prof. Jane Visvader, Prof. Gordon Smyth

CHARACTERIZING CHANGES IN EPIGENOME AND TRANSCRIPTOME DURING MAMMARY STEM CELL DIFFERENTIATION

ABSTRACT

Understanding the molecular mechanisms regulating mammary stem cell differentiation is not only important for elucidating this important biological process, but also important for discovering genes that may contribute to the initiation and progression of diseases relevant to this process such as breast cancer.

We have performed a next-generation sequencing experiment to profile the global changes in histone marks including H3K4me3 and H3K27me, when mammary stem cells differentiate into luminal progenitors and mature luminal cells. We have also carried out a whole genome expression microarray experiment to measure the gene expression changes during this differentiation process. We have developed a novel bioinformatics method to characterize the changes of histone marks genes and to correlate them with global gene expression changes.

Results: We found that both H3K4me3 and H3K27me3 played important roles in regulating mammary stem differentiation. H3K4me3 was found to induce gene expression during differentiation, ie. genes that were turned on acquired higher level of H3K4me3 marks and

genes that were turned off had reduced level of H3K4me3. However, H3K27me3 was found to have a opposite role, ie. repressing gene expression during differentiation.

This study shed new lights on how the differentiation of mammary stem cells were regulated at the molecular level. The bioinformatics analysis strategy we developed in this study has been demonstrated to be very successful in charactering global changes in the epigenome and transcriptome and in the correlations between them.

POSTER 92

Dr Susan Corley, Mr Cesar Canales, Professor Edna Hardeman, Dr Stephen Palmer, Professor Marc Wilkins

RNA-SEQ ANALYSIS OF LIP TISSUE FROM A MOUSE MODEL OF WILLIAMS-BEUREN SYNDROME

ABSTRACT

Williams-Beuren Syndrome (WBS) is a genetic disorder associated with multisystemic abnormalities, including craniofacial dysmorphology and cognitive defects. It is caused by a hemizygous microdeletion involving up to 28 genes in chromosome 7q11.23. Genotype/phenotype analysis of atypical microdeletions implicates two evolutionary-related transcription factors, GTF2I and GTF2IRD1, as prime candidates for the cause of the facial dysmorphology. Using a targeted *Gtf2ird1* knockout mouse, we have employed massively-parallel sequencing of mRNA (RNA-Seq) to understand changes in the transcriptional landscape associated with absence of *Gtf2ird1*. This in turn provides insight into the proteins likely to be activated or repressed by this transcriptional regulator. RNA-Seq data (100 bp paired-end reads) from lip tissue of 3 KO mice and 3 controls were mapped to the mouse genome (mm10) with TopHat2 and differential expression analysis was performed using edgeR and DESeq2. Our results show that the absence of GTF2IRD1 is associated with increased expression of genes involved in cellular proliferation, including growth factors, consistent with the observed phenotype of extreme thickening of the epidermal layer. We also found a decrease in expression of genes involved in certain signalling pathways including Wnt signalling, indicating dysregulation in the complex signalling network necessary for epidermal differentiation and facial skin patterning.

POSTER 93

Amanda Chamberlain, Christy Vander Jagt, Michael Goddard, Leah Marrett, Ben Hayes

DISCOVERY OF ALLELE SPECIFIC EXPRESSION IN RNASEQ DATA FROM EIGHTEEN BOVINE TISSUES

ABSTRACT

A common finding in genome wide association studies in humans, cattle and other species is that many significant hits fall outside genes. One possibility is that regulatory variants underpin a large proportion of complex trait variation. Detecting allele specific expression (ASE) is one potential method for finding regulatory variants. In order to investigate the frequency, and consistency across tissues, of ASE in cattle, we sequenced the RNA of 18 tissues from a lactating cow, collected at a single point in time. To correct for mapping bias, reads were mapped to parental genomes and then a chi squared test was performed to test for ASE at known heterozygous SNP within exons and therefore genes. Results confirmed 4 genes to be imprinted in cattle and discovered 13 genes previously found to be imprinted in mice or humans to be imprinted in cattle. After removing these imprinted genes 41,805 SNP were tested within 11,568 genes for ASE. 68% of these genes showed significant ASE in at least one tissue, within any one of the tissues 9-43% of genes showed significant ASE. The major allele expressed varied from ~55% to 100% of transcripts. 84% of the genes showing ASE did so in a subset of tissues. In general, the results suggest there are many genes in cattle that could be under cis regulatory control. This dataset provides information that will be used to find causative variants in these regulatory regions. Such causative variants would then be used to improve the accuracy of genomic selection in cattle.

POSTER 94

Dr Sonika Tyagi

STRANDED VERSUS NON-STRANDED RNA-SEQ

ABSTRACT

RNA-seq has proven to be a powerful technique for transcriptome analysis and has changed the way in which transcription can be viewed. RNA-seq can capture both sequence and abundance of the RNA species in a transcriptome, which makes it an attractive approach to study structure of the transcripts, splice variants and transcriptional activities in non-coding, intergenic or untraslated regions. Following recent proliferation of RNA-seq sample preparation protocols, there is little known about effect of various library protocols on the downstream analysis is incomplete.

In this paper we will discuss some of the common library preparation biases their effects on downstream analysis. We will be specifically comparing stranded and unstranded RNA-seq library preparation methods with respect to whole transcriptome analysis.

THANK YOU TO OUR GENEROUS SPONSORS

The organising committee for the 2014 AGTA Conference would like to take the opportunity to thank the following organisations for their valuable contribution, and ask that all delegates please consider these organisations for future business relationships.

Gold Sponsors

illumina®

Illumina is advancing human health by unlocking the power of the genome; from genome-wide discovery through to targeted validation and beyond.

At Illumina, our goal is to apply innovative technologies and revolutionary assays to the analysis of genetic variation and function, making studies possible that were not even imaginable just a few years ago.

These revolutionary tools for DNA, RNA and protein analysis are enabling rapid advances in disease research, drug development and the development of molecular tests in the clinic.

www.illumina.com

PACIFIC BIOSCIENCES™

Pacific Biosciences has developed a novel approach to studying the synthesis and regulation of DNA, RNA and proteins. Combining recent advances in nanofabrication, biochemistry, molecular biology, surface chemistry and optics, Pacific Biosciences has created a powerful technology platform called single molecule, real-time, or SMRT®, technology.

SMRT technology enables real-time analysis of biomolecules with single molecule resolution, which has the potential to transform the understanding of biological systems by providing a window into these systems that has not previously been open for scientific study.

www.pacificbiosciences.com

millennium science

Millennium Science is a specialist distribution company established in 1999 specifically to source, supply and support high technology solutions to the Australasian Life Science market. The company is well established as a leading distributor of instrumentation and reagents for the research market with recognition of performance and business practice excellence by the global companies represented, including Fluidigm and Pacific Biosciences.

www.mscience.com.au

Fluidigm

Fluidigm develops and manufactures Integrated Fluidic Circuit (IFC) technology that significantly improves productivity in life science research. Fluidigm IFCs enable the simultaneous performance of thousands of sophisticated biochemical measurements in nanolitre volumes. These systems are used for applications ranging from single-cell genomics, targeted resequencing, gene expression and digital PCR.

www.fluidigm.com

Bronze Sponsor

bio-strategy
delivering technology

Bio-Strategy is a team of highly qualified people, throughout Australia and New Zealand, who share a passion for Science.

Representing world leading suppliers, Bio-Strategy brings our customers a complete range of solutions, from basic laboratory instrumentation through to advanced technology, covering applications in the Research, Healthcare, Biotechnology and Industrial Markets.

Our focus and commitment to our customers and suppliers, is key to Bio-Strategy's excellence and success.

www.bio-strategy.com

Conference Supporters

Genome Biology

Genome Biology publishes research articles, new methods and software tools, in addition to reviews and opinions, from the full spectrum of biology, including molecular, cellular, organism or population biology studied from a genomic perspective, as well as sequence analysis, bioinformatics, proteomics, comparative biology and evolution.

www.genomebiology.com

Peter Mac

www.petermac.org

QFAB

www.qfab.org

COMBINE

www.combine.org.au

THANK YOU TO OUR TRADE EXHIBITORS

TRADE BOOTH 1

illumina®

Illumina is advancing human health by unlocking the power of the genome; from genome-wide discovery through to targeted validation and beyond.

At Illumina, our goal is to apply innovative technologies and revolutionary assays to the analysis of genetic variation and function, making studies possible that were not even imaginable just a few years ago.

These revolutionary tools for DNA, RNA and protein analysis are enabling rapid advances in disease research, drug development and the development of molecular tests in the clinic.

www.illumina.com

TRADE BOOTH 2

agrf Australian Genome Research Facility Ltd

Australian Genome Research Facility Ltd (AGRF), a not for profit company, is Australia's largest provider of genomics services and solutions. AGRF has laboratories in Brisbane, Sydney, Melbourne, Adelaide and Perth, each providing a gateway to a national network of state-of-the-art facilities, technology and expertise. From microbes to plants, animals and humans, AGRF provides a full range of services across the entire biological spectrum. We provide services to academia and industry with clients from bioscience, environmental science, biomedicine and agricultural biotechnology. Our range of services is unique in Australia and recognised internationally. AGRF is accredited by NATA to the ISO/IEC17025:2005 Quality Standard.

www.agrf.org.au

TRADE BOOTH 3

ThermoFisher SCIENTIFIC

Thermo Fisher Scientific is the world leader in serving science, enabling our customers to make the world healthier, cleaner and safer. We are the leading provider of analytical instruments, equipment, reagents and consumables, software and services for research, analysis, discovery and diagnostics. In Australia and New Zealand, you can access the renowned Thermo Scientific portfolio as well as premium, through Thermo Fisher Scientific. Our comprehensive range includes high-end analytical and process instrumentation, laboratory equipment, a complete range of consumables, chemicals and reagents as well as the service and support to optimise your business.

www.thermofisher.com.au

TRADE BOOTH 4

life technologies™

Life Technologies™ products harness the power of science to transform lives. As a member of the Thermo Fisher Scientific family of brands, our instruments, everyday tools, and services offer high-quality, innovative life science solutions for every lab and application.

www.lifetechnologies.com

TRADE BOOTH 5

TECAN.

Tecan is a leading global provider of laboratory instruments and solutions in biopharmaceuticals, forensics and clinical diagnostics. We specialize in the development, production and distribution of automated workflow solutions for laboratories in the life sciences sector.

We offer expandable, reconfigurable solutions that cover all your genomic automation needs including DNA extraction, Normalization, PCR set up, DNA shearing and NGS sample preparation. For example, assay-specific set up for PCR, SNP scoring, sequencing, reaction clean up, as well as library fragment preparation and emPCR enrichment for next generation on the same platform makes the Tecan Freedom EVO a truly powerful tool for the genomics laboratory.

www.tecan.com

TRADE BOOTH 6

Roche

Roche offers research solutions, from nucleic acid extraction and qPCR through to targeted enrichment using NimbleGen SeqCap reagents.

Roche is committed long-term to sequencing, recently acquiring Genia Technologies and investing in Stratos Genomics, both for single molecule sequencing nanopore technologies; the goal being higher throughput, short turnaround and reduced costs.

www.lifescience.roche.com

TRADE BOOTH 7

Agilent Technologies

Agilent is a global leader in target enrichment solutions for next-generation sequencing and genomic microarrays. By synthesizing custom complex mixtures of long oligonucleotides, SureSelect and HaloPlex target enrichment solutions enable researchers to identify genomic regions of interest for focused, cost-effective variant profiling, with workflows that generate sequencing-ready libraries in just one day. For more information visit Agilent Genomics.

www.genomics.agilent.com

TRADE BOOTH 8 & 9

millennium science | Fluidigm | PACIFIC BIOSCIENCES

Millennium Science is a specialist distribution company established in 1999 specifically to source, supply and support high technology solutions to the Australasian Life Science market. The company is well established as a leading distributor of instrumentation and reagents for the research market with recognition of performance and business practice excellence by the global companies represented, including Fluidigm and Pacific Biosciences.

Pacific Biosciences has developed a novel approach to studying the synthesis and regulation of DNA, RNA and proteins. Combining recent advances in nanofabrication, biochemistry, molecular biology, surface chemistry and optics, Pacific Biosciences has created a powerful technology platform called single molecule, real-time, or SMRT®, technology.

SMRT technology enables real-time analysis of biomolecules with single molecule resolution, which has the potential to transform the understanding of biological systems by providing a window into these systems that has not previously been open for scientific study.

Fluidigm develops and manufactures Integrated Fluidic Circuit (IFC) technology that significantly improves productivity in life science research. Fluidigm IFCs enable the simultaneous performance of thousands of sophisticated biochemical measurements in nanolitre volumes. These systems are used for applications ranging from single-cell genomics, targeted resequencing, gene expression and digital PCR

www.mscience.com.au

www.fluidigm.com

www.pacificbiosciences.com

TRADE BOOTH 10

华大科技 BGI tech

To be your premier scientific partner, BGI Tech is committed to providing high quality and innovative products and solutions in biological, agricultural and medical research.

BGI Tech was founded in 2012 and is headquartered in Shenzhen, China. Our service network covers more than 100 countries and regions. Currently, BGI Tech collaborates with over 10,000 organizations and 30,000 partners, providing highly reliable services for thousands of research projects, including several large-scale international genome projects. To date, BGI Tech has contributed to over 600 publications, with more than 100 articles appearing in top-tier journals such as Nature and Science.

www.bgitechsolutions.com

TRADE BOOTH 11

Ramaciotti Centre for Genomics

The Ramaciotti Centre is a leading genomic service provider located at the University of New South Wales, Sydney. Our services include next-generation sequencing, Sanger sequencing, microarray, SNP genotyping, CNV analysis and gene expression analysis including single cell gene expression studies.

We produce data of the highest quality and provide excellent personalised customer service. Our professional team has many years of experience facilitating projects from design through to downstream analysis.

We are a CSPro certified provider of Illumina sequencing and an Authorised Affymetrix Service Provider.

www.ramaciotti.unsw.edu.au

TRADE BOOTH 12

GENESEARCH

Genesearch is the Australian distributor of New England Biolabs products, including the NEBNext product range for NGS library preparation. NEBNext reagents provide solutions for the major NGS platforms. Featuring NEB's NGS-optimized polymerase, Q5, for robust performance with a wide range of amplicons, regardless of GC content. Please visit the NEB website to see the latest products including 96 Dual Index Primers compatible with the Illumina platform.

www.genesearch.com.au

TRADE BOOTH 13

bio-strategy
delivering technology

Bio-Strategy is a team of highly qualified people, throughout Australia and New Zealand, who share a passion for Science.

Representing world leading suppliers, Bio-Strategy brings our customers a complete range of solutions, from basic laboratory instrumentation through to advanced technology, covering applications in the Research, Healthcare, Biotechnology and Industrial Markets.

Our focus and commitment to our customers and suppliers, is key to Bio-Strategy's excellence and success.

www.bio-strategy.com

TRADE BOOTH 14

QIAGEN

QIAGEN is a global leader in sample and assay technologies for life sciences, applied testing and molecular diagnostics. Its products are considered standards that comprise complete solutions from sample to result. QIAGEN offers the broadest portfolio of molecular diagnostic assays for infectious diseases including the only test for human papillomavirus (HPV), which has both FDA and CE-approvals. The company has developed a comprehensive portfolio of more than 500 proprietary, consumable products and instruments for sample collection, nucleic acid and protein handling, separation, purification, detection, and open and target specific assays.

www.qiagen.com

TRADE BOOTH 15

SARSTEDT

Sarstedt Australia is recognised in Australia and internationally as a quality manufacturer of a comprehensive range of diagnostic, laboratory and medical products that includes both equipment and consumables for use in pathology, laboratory, research, transfusion medicine, blood banking and hospital/ incontinence applications. Sarstedt's product range includes plastic disposables for the sophisticated requirements of cell culture research and PCR applications. The range includes Serological Pipettes, Tissue Culture Flasks, TC Dishes and TC Multi Well Plates, Filters, PCR Plates, PCR Strips and Cups, Adhesive Seals, Cryo Preservation Tubes, as well as a comprehensive range of general purpose laboratory tubes and plastic consumables.

www.sarstedt.com

TRADE BOOTH 16

TrendBio

TrendBio is a leading supplier of specialised instrumentation for the Life Sciences. During the AGTA 2014 meeting we will be promoting;

- DNA Shearing (Covaris)
- FFPE Extraction (DNA/RNA)
- Chromatin Shearing
- RNA and DNA Analysis (capillary gel electrophoresis)
- Microscale Thermophoresis – Biomolecule Interactions (NanoTemper)
- Automated Western Blotting (Protein Simple)
- Pre-Clinical Imaging

www.trendbio.com.au

TRADE BOOTH 17

mimr-phi institute

The MHTP Medical Genomics Facility at the Monash Health Translation Precinct (MHTP) in Victoria uses the latest and most innovative genomic technologies to service research institutes throughout Australia. We maintain an excellent reputation for provision of the highest quality genomic data and client support.

Our not-for-profit, NATA-accredited facility offers Sanger and Next Generation Sequencing, High Content Screening, Microarray, Real-Time qPCR and Fluidigm Single Cell technologies. Combined with our professional consultation, we aim to facilitate and accelerate genomic-based translational research.

www.mimr-phi.org

TRADE BOOTH 18

OnQ software

OnQ Software is dedicated in providing leading edge custom technology solutions with the aim of enabling end to end integration across your entire workplace using our suite of state of the art software, tools and highly trained experts who can tailor a solution for your business needs.

www.onqsoft.com.au

TRADE BOOTH 19

IDT INTEGRATED DNA TECHNOLOGIES

Integrated DNA Technologies (IDT) is a leader in manufacturing and developing products for the research and diagnostic life science market. IDT is also the world leader in nucleic acid synthesis and serves the areas of academic research, biotechnology, and pharmaceutical development.

Product offerings include DNA and RNA oligos, qPCR assays and probes, siRNA duplexes, DNA fragments and custom gene synthesis. Pioneering methods in oligo synthesis and quality control allow IDT to be able to synthesize DNA oligos up to 2kb. IDT products support applications such as DNA sequencing, DNA amplification, SNP detection, microarray analysis, expression profiling, gene quantification, and functional genomics.

www.idtdna.com

TRADE BOOTH 20

scientifix LIFE

Scientifix is an Australian-owned distribution company established in 1998. Our product lines include Macherey Nagel German-quality extraction kits for all your DNA/RNA and clean-up purification requirements. Macherey Nagel isolation products are also formatted and optimised for seamless automation on a number of leading robotic platforms. Macherey Nagel also offers magnetic particles for reaction clean-up and size selection in NGS library construction.

Clontech Laboratories is a leading supplier for innovative kits and reagents for gene discovery, regulation, and function. Clontech's SMARTer cDNA synthesis is the industry standard for exceptional sensitivity and complete transcriptome coverage as it synthesizes accurate, full-length cDNA directly from cells or total RNA. SMARTer Ultra Low Input RNA Kit for Sequencing – v3 is the gold standard in cDNA Synthesis for Single-Cell Transcriptome Analysis and compatible with leading NGS platforms.

Takara Bio has positioned itself as a leading manufacturing company in the Japanese life science market. Takara Bio's R&D scientists are expert enzymologists who are continually working to improve enzymes and buffers for PCR, real-time PCR, and RT-PCR.

www.scientifix.com.au

TRADE BOOTH 21

eppendorf

Eppendorf is a leading life science company that develops and sells instruments, consumables, and services for Liquid, Sample and Cell handling in laboratories worldwide.

Eppendorf's product range includes pipettes and automated pipetting systems, dispensers, centrifuges, mixers, spectrometers and DNA amplification equipment as well as ultra-low temperature freezers, platform shakers, CO2 Incubators, Bioprocess equipment and cell manipulation systems. Associated consumables like pipette tips, test tubes, microtiter plates, cell culture consumables and disposable bioreactors vessels complement the instruments for the highest quality workflow solutions.

Our mission is to be a synonym for customer-focused processes, innovative technologies, premium products and services to improve human living conditions.

www.eppendorf.com.au

TRADE BOOTH 22

SiSYSTEMS

SI Systems provides IT services to Life Science Markets and specialises in enterprise IT Infrastructure solutions for Big data storage, custom software development and Bioinformatics data analysis services. We can help you choose the right data storage solutions ranging from 10TB to hundreds of PB (EMC: Isilon,VNX, Pivoltal or Coraid: Flash, Deep storage), along with CISCO and Dell compute systems. Contact us for expert Bioinformatic solutions for analysis and biological interpretation of high-throughput sequencing data.

www.sisystems.com.au

TRADE BOTH 23

PerkinElmer
For the Better

PerkinElmer is a premier provider of cutting-edge technologies and solutions enabling researchers to create more efficient and quicker life-saving and enhancing medicines and diagnostic tests. As a leading solutions innovator, PerkinElmer bridges the gap between in vitro assays and in vivo results, allowing researchers to translate their results into cures for human disease.

Our wide range of systems consist of instrumentation for automation and detection solutions, cellular imaging and analysis hardware and software, and an industry-leading portfolio of consumables products. Our drug discovery and research reagents support our core expertise in cellular sciences, time-resolved fluorescence, chemiluminescence, radioactive labelling, and the detection of proteins and nucleic acids.

The company's offerings also include state-of-the-art microfluidics, lab automation & liquid handling, tissue microscopy, preclinical imaging technologies, and discovery & development outsourcing solution.

www.perkinelmer.com

TRADE BOTH 24

DNA genotek

DNA Genotek provides high-quality biological sample collection, stabilization and preparation products for human genetics, microbiology and animal genetics. The company's products protect and stabilize multiple sample types for long-term storage at ambient temperature to ensure the highest quality results for genetic analysis and testing. Due to their reliability and ease-of-use, DNA Genotek's products are used by thousands of academic, biotechnology, diagnostic, agriculture, and other leading institutions around the globe.

www.dnagenotek.com

TRADE BOOTH 25

BIO-RAD

Bio-Rad Laboratories, Inc. Designs, manufactures, and distributes a broad range of innovative products and solutions for the life science research and clinical diagnostic markets. Bio-Rad is renowned for its commitment to quality and customer service among university and research institutions, hospitals, public health and commercial laboratories, as well as the biotechnology, pharmaceutical, and food safety industries. Founded in 1952, Bio-Rad is based in California, and serves more than 100,000 research and healthcare industry customers globally.

www.bio-rad.com

TRADE BOOTH 26

MACROGEN
Advancing through Genomics

Macrogen is the current market leader in biotechnology services based in Korea, providing pivotal research tools to a global client base. The company provides high quality genomic sequencing, microarray, oligo synthesis, knock-out mouse, and bioinformatics services at competitive prices. Using only the most up to date technology available on the market, Macrogen is able to facilitate large scale evolutionary studies, whilst still taking care to provide smaller scale investigations, customised to the client's needs.

www.macrogen.com

TRADE BOOTH 27

GeneWorks

GeneWorks is an Adelaide-based genomic services provider and supplier of innovative products for molecular research. Supported technologies include MassARRAY from Agena Bioscience, PCR reagents and NGS kits from KAPA Biosystems, SAGE Science Pippin Prep and Blue Pippin for library preparation, Exiqon miRNA solutions, and KASP genotyping from LGC. Genetic analysis services include SNP genotyping assays (custom designed and Agena panels). GeneWorks now also offers a sample validation service using Agena's Exome ID panel.

www.geneworks.com.au

EXHIBITION FLOOR PLAN

EXHIBITOR	BOOTH
Illumina	1
AGRF	2
Thermo Fisher SCIENTIFIC	3
Life Technologies	4
Tecan Australia	5
Roche	6
Agilent Technologies	7
Millennium Science	8&9
BGI Tech	10
Ramaciotti Centre for Genomics	11
GENESEARCH	12
Bio-Strategy	13
QIAGEN	14
SARSTEDT	15
TrendBio	16
MIMR - PHI Institute (Monash Health)	17
OnQ Software	18
Intergrated DNA	19
Scientifix LIFE	20
Eppendorf	21
SI Systems	22
PerkinElmer	23
DNA Genotek	24
Bio-Rad	25
Macrogen	26
GeneWorks	27

AGTA 2014 DELEGATE LIST

FIRST NAME	LAST NAME	ORGANISATION	STATE	COUNTRY
Matias	Abregu	PerkinElmer	VIC	Australia
Satoru	Akama	Tokyo Institute Of Technology		Japan
Raymond	Ali	University Of Tasmania	TAS	Australia
Richard	Allcock	University Of Western Australia	WA	Australia
Janelle	Allen	Illumina Australia Pty Ltd	VIC	Australia
Mohammed	Alnakhli	Flinders University	SA	Australia
Mohammed	Alqahtani	CEGMR		Saudi Arabia
Mary	Anderson	SydPath, St Vincent's Hospital	NSW	Australia
Stuart	Archer	Monash University	VIC	Australia
Josh	Ardley	Tecan Australia	VIC	Australia
Nicola	Armstrong	University Of Sydney	NSW	Australia
Marie-Liesse	Asselin-Labat	The Walter & Eliza Hall Institute Of Medical Research	VIC	Australia
Kelly	Atkinson	University Of Auckland		New Zealand
Melanie	Bahlo	The Walter & Eliza Hall Institute Of Medical Research	VIC	Australia
Gia	Bawa	Geneworks Pty Ltd	SA	Australia
Anthony	Beckhouse	Bio-Rad Labratories	NSW	Australia
Alex	Beesley	Telethon Kids Institute	WA	Australia
Katrina	Bell	Murdoch Childrens Research Institute	VIC	Australia
Nigel	Bennett	University Of Queensland	QLD	Australia
Marnie	Blewitt	The Walter & Eliza Hall Institute Of Medical Research	VIC	Australia
Justin	Borevitz	Australian National University	ACT	Australia
David	Bowtell	Peter MacCallum Cancer Centre	VIC	Australia
Titus	Brown	Department of Microbiology and Molecular Genetics		USA
Magdalena	Budzinska	Centenary Institute	NSW	Australia
Suzanne	Butcher	Australian Institute For Bioengineering And Nanotechnology	QLD	Australia
Keith	Byron	Healthscope	VIC	Australia
Sophia	Cameron-Christie	University Of Otago		New Zealand
Peter	Campbell	Wellcome Trust Sanger Institute		United Kingdom
Minh Duc	Cao	The University Of Queensland	QLD	Australia
Karen	Carpenter	PathWest Diagnostic Genomics	WA	Australia
Kim	Carter	Telethon Kids Institute	WA	Australia

FIRST NAME	LAST NAME	ORGANISATION	STATE	COUNTRY
Daniel	Catchpoole	Biospecimens Research And Tumour Bank, Children's Cancer Research Unit	NSW	Australia
Alistair	Chalk	St Vincents Institute	VIC	Australia
Amanda	Chamberlain	Department Of Environment And Primary Industries	VIC	Australia
Anandhakumar	Chandran	Kyoto University		Japan
Jac	Charlesworth	University Of Tasmania	TAS	Australia
Rima	Chaudhuri	Garvan Institute Of Medical Research	NSW	Australia
Yunshun	Chen	The Walter & Eliza Hall Institute Of Medical Research	VIC	Australia
Wenhan	Chen	University Of Queensland	QLD	Australia
Tyrone	Chen	University Of Queensland	QLD	Australia
Belinda	Chong	Victorian Clinical Genetics Services	VIC	Australia
Cheuk Wang	Chung	National University Of Singapore		Singapore
Nicole	Cloonan	QIMR Berghofer Medical Research Institute	QLD	Australia
Ana	Conesa	Genomics of Gene Expression Lab		Spain
Thomas	Conway	IBM Research Australia	VIC	Australia
Vincent	Corbin	The Walter & Eliza Hall Institute Of Medical Research	VIC	Australia
Susan	Corley	SBI, University Of NSW	NSW	Australia
Mark	Cowley	Garvan Institute Of Medical Research	NSW	Australia
Jeffrey	Craig	Murdoch Childrens Research Institute	VIC	Australia
Mark	Crowe	QFAB Bioinformatics	QLD	Australia
Tamsyn	Crowley	Deakin University	VIC	Australia
Mark	Cruickshank	Telethon Kids Institute	WA	Australia
Helen	Cumming	Mimr-Phi	VIC	Australia
Robert	Curtain	Bio-Rad Laboratories	NSW	Australia
Patrick	Danoy	Roche Diagnostics	NSW	Australia
Aaron	Darling	University Of Technology Sydney	NSW	Australia
Harriet	Dashnow	Murdoch Childrens Research Institute	VIC	Australia
Nadia	Davidson	Murdoch Childrens Research Institute	VIC	Australia
Mark	Davis	PathWest Diagnostic Genomics	WA	Australia
Sarah-Jane	Dawson	Peter MacCallum Cancer Centre	VIC	Australia
Niantao	Deng	Garvan Institute Of Medical Research	NSW	Australia
Leanne	Dibbens	Sansom Institute For Health Research, School Of Pharmacy And Medical Sciences, U	SA	Australia
Alison	Digney	Life Technologies	VIC	Australia
Marcel	Dinger	Garvan Institute Of Medical Research	NSW	Australia
James	Doecke	CSIRO	QLD	Australia
Kenneth	Doig	Peter MacCallum Cancer Centre	VIC	Australia
Maria	Doyle	Peter MacCallum Cancer Centre	VIC	Australia

FIRST NAME	LAST NAME	ORGANISATION	STATE	COUNTRY
Desiree	Du Sart	Victorian Clinical Genetics Services	VIC	Australia
Bjorn	Dueholm	University Of South Australia	SA	Australia
Ankit	Dutta	The University Of Adelaide	SA	Australia
Jason	Ellul	Peter MacCallum Cancer Centre	VIC	Australia
Michael	Erlichster	The Walter & Eliza Hall Institute Of Medical Research	VIC	Australia
Kelly	Ewen-White	Life Technologies	VIC	Australia
Miriam	Fanjul Fernandez	Murdoch Childrens Research Institute	VIC	Australia
Cassy	Faux	Australian Antarctic Division	TAS	Australia
Andrew	Fellowes	Peter MacCallum Cancer Centre	VIC	Australia
Christoffer	Flensburg	McRi	VIC	Australia
Cathy	Ford	DNA Genotek		Canada
Sue	Forrest	Australian Genome Research Facility	VIC	Australia
Archa	Fox	The Harry Perkins Institute of Medical Research, UWA	WA	Australia
Iain	Fraser	National Institute of Health		USA
Michelle	French	AgResearch Ltd		New Zealand
Saskia	Freytag	The Walter & Eliza Hall Institute Of Medical Research	VIC	Australia
Melissa	Fullwood	Cancer Science Institute Singapore		Singapore
Devika	Ganesamoorthy	Institute For Molecular Biosciences	QLD	Australia
Jonathan	Gannoulis	OnQ Software	VIC	Australia
Amee	George	Peter MacCallum Cancer Centre	VIC	Australia
Brian	Gloss	Kinghorn Cancer Centre	NSW	Australia
Anita	Goldinger	University Of Queensland	QLD	Australia
David	Goode	Peter MacCallum Cancer Centre	VIC	Australia
Jodee	Gould	MHTP, Medical Genomics Facility	VIC	Australia
Alexander	Gout	Telethon Kids Institute	WA	Australia
Parry	Guilford	University of Otago		New Zealand
Richard	Harrison	Bio-Rad Laboratories	NSW	Australia
Julie	Haste	Agilent Technologies	VIC	Australia
Jean Christophe	Hatinguais	Kapa Biosystems		USA
Lyndal	Henden	The Walter & Eliza Hall Institute Of Medical Research	VIC	Australia
Duncan	Hewett	University Of Adelaide	SA	Australia
Sarah	Hickey	Life Technologies	VIC	Australia
Tony	Hoady	Eppendorf	NSW	Australia
Aliaksei	Holik	The Walter & Eliza Hall Institute Of Medical Research	VIC	Australia
Kathryn	Holt	University of Melbourne	VIC	Australia
Katharine	Howell	ARC CoE In Plant Energy Biology, University Of Western Australia	WA	Australia

FIRST NAME	LAST NAME	ORGANISATION	STATE	COUNTRY
Michelle	Howie	Roche Diagnostics	NSW	Australia
Ji-Fan (Sarah)	Hsieh	The Australian National University	ACT	Australia
Siobhan	Hughes	King's College London		United Kingdom
Sally	Hunter	Peter MacCallum Cancer Centre	VIC	Australia
Gyorgy	Hutvagner	UTS	NSW	Australia
Michael	Inouye	University Of Melbourne	VIC	Australia
Kaushala	Jayawardana	University Of Sydney	NSW	Australia
Aaron	Jeffs	University Of Otago		New Zealand
Margaret	Jordan	Comparative Genomics Centre	QLD	Australia
Chol-Hee	Jung	VLSCI Life Sciences Computation Centre	VIC	Australia
David	Kainer	Australian National University	ACT	Australia
Naga	Kasinadhuni	AGRF	QLD	Australia
Gurpreet	Kaur	Millennium Science	VIC	Australia
Anabel	Kearney	Liverpool Hospital	NSW	Australia
Farzaneh	Kordbacheh	University Of Queensland	QLD	Australia
Carsten	Kulheim	Australian National University	ACT	Australia
Paul	Lacaze	Millennium Science	VIC	Australia
Nicole	Lake	Murdoch Childrens Research Institute	VIC	Australia
Timo	Lassmann	Telethon Kids Institute	WA	Australia
Charity	Law	University Of Zurich		Switzerland
David	Lawrence	ACRF Cancer Genomics Facility, Centre For Cancer Biology	SA	Australia
Tim	Lawrence	NZGL/Auckland University		New Zealand
Lien	Le	Bioinformatics Resource Australia Of EMBL Australia (BRAEMBL)	QLD	Australia
Marcus	Lefebure	Peter MacCallum Cancer Centre	VIC	Australia
Xuan	Li	BGI Tech Solutions Co. Ltd		China
Xi	Li	CSIRO	ACT	Australia
Jason	Li	Peter MacCallum Cancer Centre	VIC	Australia
Webber W.P.	Liao	Bioinformatics Resource EMBL Australia; NCI Specialised Facility In Bioinformati	QLD	Australia
Ruby	Lin	University Of New South Wales	NSW	Australia
Tao	Liu	Children's Cancer Institute Australia	NSW	Australia
Jianhua	Liu	Institute For Comprehensive Utilisation Of Marine Biological Resources		China
Sheng	Liu	The Walter & Eliza Hall Institute Of Medical Research	VIC	Australia
Andrew	Lonsdale	The University Of Melbourne	VIC	Australia
Evangelina	López De Maturana	Spanish National Cancer Research Centre		Spain

FIRST NAME	LAST NAME	ORGANISATION	STATE	COUNTRY
David	Lovell	CSIRO/Australian Bioinformatics Network	ACT	Australia
Robyn	Lukeis	SydPath, St. Vincent's Hospital	NSW	Australia
Vesna	Lukic	The Walter & Eliza Hall Institute Of Medical Research	VIC	Australia
Aaron	Lun	The Walter & Eliza Hall Institute Of Medical Research	VIC	Australia
Richard	Lupat	Peter MacCallum Cancer Centre	VIC	Australia
Donia	Macartney-Coxson	Institute Of Environmental Science And Research		New Zealand
Ruth	MacKinnon	Victorian Cancer Cytogenetics Service	VIC	Australia
Paul	Maclean	AgResearch Ltd		New Zealand
Gregory	Maes	James Cook University	QLD	Australia
Ian	Majewski	The Walter & Eliza Hall Institute Of Medical Research	VIC	Australia
Jovana	Maksimovic	Murdoch Childrens Research Institute	VIC	Australia
Nauman	Maqbool	AgResearch Ltd		New Zealand
Vikki	Marshall	Australasian Genomics Technologies Association	QLD	Australia
Justine	Marum	SA Pathology	SA	Australia
Elizabeth	Mason	University Of Queensland	QLD	Australia
Karen	Mather	University Of New South Wales	NSW	Australia
John	Mattick	Garvan Institute of Medical Research	NSW	Australia
Damara	McAndrew	Genesearch	QLD	Australia
Mark	McCabe	Garvan Institute Of Medical Research	NSW	Australia
Amy	McCart Reed	UQCCR	QLD	Australia
Ann	McCormack	Garvan Institute Of Medical Research	NSW	Australia
Helen	McCormick	Victor Chang Cardiac Research Institute	NSW	Australia
George	McGillivray	Murdoch Childrens Research Institute	VIC	Australia
Liam	McIntyre	SA Pathology	SA	Australia
Cliff	Meldrum	Hunter New England Health	NSW	Australia
Rebecca	Miller	Integrated DNA Technologies Pte	NSW	Australia
Gisela	Mir	Peter MacCallum Cancer Centre	VIC	Australia
Mahtab	Mirmomeni	IBM Research Australia	VIC	Australia
Mika	Mitropoulos	Thermo Fisher Scientific	VIC	Australia
Cath	Moore	QIAGEN	VIC	Australia
Ryan	Morin	Simon Fraser University		Canada
Hayley	Mountford	Murdoch Childrens Research Institute	VIC	Australia
Kevin	Murray	Australian National University	ACT	Australia
Alexander	Myburg	University of Pretoria		South Africa
Stacey	Nelson	Bio-Strategy		NZ
Kathryn	North	Murdoch Childrens Research Institute	VIC	Australia

FIRST NAME	LAST NAME	ORGANISATION	STATE	COUNTRY
Harald	Oey	La Trobe University	VIC	Australia
Benjamin	Ong	Murdoch Childrens Research Institute	VIC	Australia
Alicia	Oshlack	Murdoch Childrens Research Institute	VIC	Australia
Yalchin	Oytam	CSIRO	NSW	Australia
Arunkumar	Padmanaban	Agilent Technologies	VIC	Australia
Lisa	Paiman	Healthscope	VIC	Australia
Tony	Papenfuss	The Walter & Eliza Hall Institute Of Medical Research	VIC	Australia
Dhru	Patel	OnQ Software	VIC	Australia
Ralph	Patrick	University Of Queensland	QLD	Australia
Ellis	Patrick	University Of Sydney	NSW	Australia
John	Pearson	QIMR Berghofer Medical Research Institute	QLD	Australia
Stephen	Pederson	University Of Adelaide	SA	Australia
Dilmi	Perera	Lowy Cancer Research Center, UNSW	NSW	Australia
Victoria	Perreau	The Florey	VIC	Australia
Belinda	Phipson	Murdoch Childrens Research Institute	VIC	Australia
Justin	Pickering	Sarstedt Australia	SA	Australia
Michael	Poidinger	Singapore Immunology Network		Singapore
Leanne	Purins	SA Pathology	SA	Australia
Kelly	Quek	Queensland Centre For Medical Genomics	QLD	Australia
Ashfaqur	Rahman	CSIRO	TAS	Australia
Louis	Ranjard	The University Of Auckland		New Zealand
Coralie	Reich	DEPI	VIC	Australia
Michael	Ricos	University Of South Australia	SA	Australia
Scott	Ritchie	Medical Systems Biology Lab	VIC	Australia
Matthew	Ritchie	The Walter & Eliza Hall Institute Of Medical Research	VIC	Australia
Bronwyn	Robertson	Ramaciotti CentreFor Genomics	NSW	Australia
Daniel	Roden	Garvan Institute Of Medical Research	NSW	Australia
Sandra	Romanin	Westwick-Farrow Media	NSW	Australia
Jason	Ross	CSIRO	NSW	Australia
Fernando	Rossello	Victorian Bioinformatics Consortium	VIC	Australia
Alan	Rubin	The Walter & Eliza Hall Institute Of Medical Research	VIC	Australia
Kalyan	Rudraraju	Eppendorf	NSW	Australia
Georgie	Ryland	Peter MacCallum Cancer Centre	VIC	Australia
Jafar	S. Jabbari	Australian Genome Research Facility	VIC	Australia
Simon	Sadedin	Murdoch Childrens Research Institute	VIC	Australia
Mnasri	Sameh	National Gene Bank Of Tunisia		Tunisia
Andreas	Schlitzer	Singapore Immunology Newtork		Singapore
Andreas	Schreiber	ACRF South Australian Cancer Genomics Facility	SA	Australia

FIRST NAME	LAST NAME	ORGANISATION	STATE	COUNTRY
Jan	Schroeder	The Walter & Eliza Hall Institute Of Medical Research	VIC	Australia
Mark	Schultz	University Of Melbourne	VIC	Australia
Hamish	Scott	SA Pathology	SA	Australia
Robert	Sebra	ICAHN School of Medicine		USA
Torsten	Seemann	Victorian Bioinformatics Consortium	VIC	Australia
Tim	Semple	Peter MacCallum Cancer Centre	VIC	Australia
Jeremy	Shearman	National Centre For Genetic Engineering And Biotechnology		Thailand
Wei	Shi	The Walter & Eliza Hall Institute Of Medical Research	VIC	Australia
Christopher	Sibley	University College London		United Kingdom
Beth	Signal	Garvan Institute Of Medical Research	NSW	Australia
Kaylene	Simpson	Peter MacCallum Cancer Centre	VIC	Australia
Melanie	Smith	Victorian Clinical Genetics Services	VIC	Australia
Adam	Smolenski	University Of Tasmania	TAS	Australia
Gordon	Smyth	The Walter & Eliza Hall Institute Of Medical Research	VIC	Australia
Maggie	Soliman	PerkinElmer	VIC	Australia
Helen	Speirs	Ramaciotti CentreFor Genomics	NSW	Australia
Akanksha	Srivastava	University Of Western Australia	WA	Australia
Doris	Stangl	SA Pathology	SA	Australia
Jo-Ann	Stanton	University Of Otago		New Zealand
Dorothy	Steane	University Of Tasmania & University Of The Sunshine Coast	TAS	Australia
Phil	Stephens	Foundation Medicine, Inc.		USA
Tim	Stuart	The University Of Western Australia	WA	Australia
Siamak	Tafavogh	University Of Technology Sydney	NSW	Australia
Rick	Tankard	The Walter & Eliza Hall Institute Of Medical Research	VIC	Australia
Graham	Taylor	University Of Melbourne	VIC	Australia
Luisa	Teasdale	Museum Victoria	VIC	Australia
Erik	Thompson	Queensland University of Technology	QLD	Australia
Daniel	Thomson	Garvan Institute Of Medical Research	NSW	Australia
Natalie	Thorne	The Walter & Eliza Hall Institute Of Medical Research	VIC	Australia
Amali	Thrimawithana	The New Zealand Institute For Plant & Food Research Limited		New Zealand
Peri	Tobias	University Of Sydney	NSW	Australia
Kim	Ton	University Of Otago		New Zealand
Richard	Tothill	Peter MacCallum Cancer Centre	VIC	Australia
Peter	Tsai	University Of Auckland		New Zealand
Nikki	Tsoudis	Scientifix	VIC	Australia

FIRST NAME	LAST NAME	ORGANISATION	STATE	COUNTRY
Sonika	Tyagi	The Australian Genome Research Facility	VIC	Australia
Mark	Van Der Hoek	SA Pathology	SA	Australia
Tracey	Van Stijn	AgResearch Ltd		New Zealand
Christy	Vander Jagt	Department Of Environment And Primary Industries	VIC	Australia
Vivien	Vasic	Monash Health Translation Precinct	VIC	Australia
Oliver	Vasilevski	Fluidigm Corporation		USA
Violeta	Velkoska-Ivanova	Healthscope	VIC	Australia
Matthew	Wakefield	Life Sciences Computational Centre, VLSCI	VIC	Australia
Mark	Waltham	St Vincents Institute	VIC	Australia
Paul	Wang	ACRF South Australian Cancer Genomics Facility, Centre For Cancer Biology, SA Pa	SA	Australia
Trent	Warburton	TrendBio Pty Ltd	VIC	Australia
Paul	Waring	The University Of Melbourne	VIC	Australia
Emma	Whitelaw	La Trobe University	VIC	Australia
Liam	Williams	University Of Auckland		New Zealand
Trevor	Wilson	Monash Health Translation Precinct	VIC	Australia
Barbara	Wold	California Institute of Technology		USA
Stephen	Wong	Peter MacCallum Cancer Centre	VIC	Australia
Janice	Woods	Millennium Science	VIC	Australia
James	Wynne	CSIRO Biosecurity Flagship	VIC	Australia
Monika	Zavodna	Otago Genomics & Bioinformatics Facility, Uni Of Otago		New Zealand
Liang	Zhang	Chinese National Human Genome Center At Shanghai		China
Kerong	Zhang	The Australian National University	ACT	Australia
Li	Zhou	The Tumour Bank At The Children's Hospital At Westmead	NSW	Australia

2015 AGTA Conference

AUSTRALASIAN GENOMIC TECHNOLOGIES ASSOCIATION

11-14 October

CROWNE PLAZA HUNTER VALLEY

THE 2015 AGTA CONFERENCE

will cover themes that include cancer genomics, epigenomics, bioinformatics, transcriptomics, functional genomics, clinical sequencing and plant genomics with a special focus on the genomics of good wine. We have invited an outstanding list of international and national speakers plus have engaged industry to showcase the best in new genomic technologies. Please come and join us for some great networking and cutting edge science in the heart of Australia's top food and wine country.

We look forward to seeing you at AGTA 2015 in the Hunter Valley to share the excitement of cutting-edge science!

Save the Date!

11-14 October 2015

CO-CONVENORS

Marcel Dinger
Garvan Institute of Medical Research, Sydney

Carsten Kulheim
Australian National University, Canberra

Any Questions?

Contact Brigitte at Leishman Associates
brigitte@leishman-associates.com.au
Phone: +61 3 6234 7844

Leishman
it's all about people

AGTA 2014
SPONSORS AND EXHIBITORS