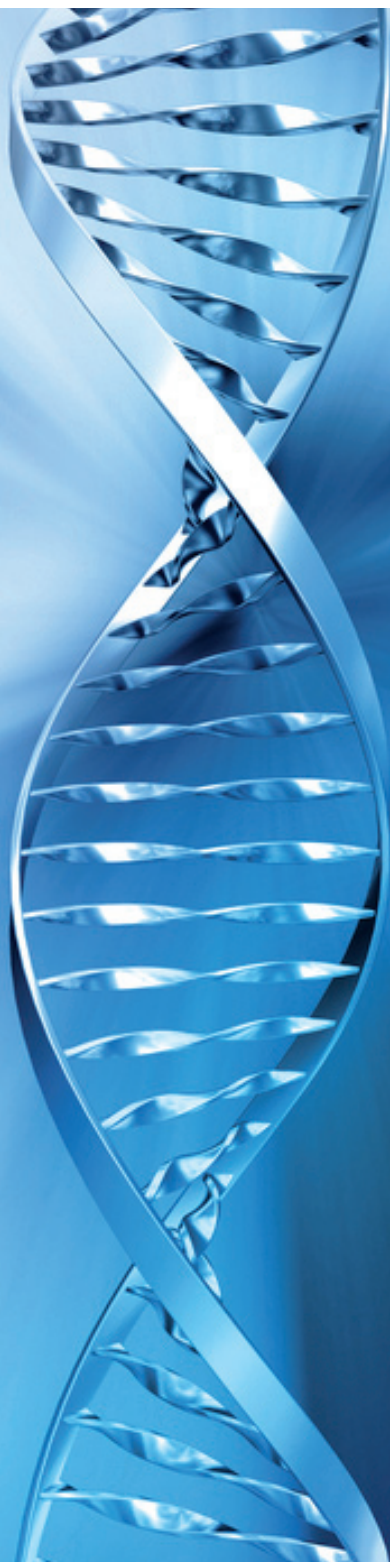


EMBL
Australia

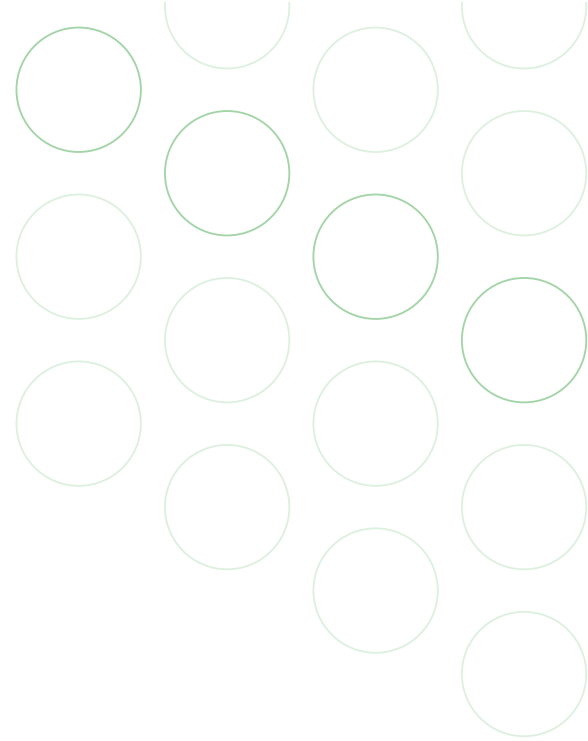


Bioinformatics for the Future

March 2009



A series of workshops and seminars sponsored by EMBL Australia and supported by the Australian Academy of Technological Sciences and Engineering (ATSE) and the Australian Government Department of Innovation, Industry, Science and Research (DIISR)



Introduction

Bioinformatics—the union of biology, computer science, and information technology —is generating many research opportunities, and many challenges.

This mix of opportunities and challenges was a constant theme throughout the series of workshops and seminars organised in conjunction with the March 2009 visit of Dr Ewan Birney, Senior Scientist of the European Bioinformatics Institute (EBI) of the EMBL.

This document summarises the key activities during Dr Birney's visit and articulates some of the key issues identified in the workshop and seminar series. Activities in each of the mainland states and the ACT enabled over 130 participants, including life scientists, government policy makers and bioinformaticians, to provide input.

I would like to take this opportunity to thank the members of the Working Group who so ably coordinated the various activities in each city and helped to make this the successful event that it was.

We all owe a debt of gratitude to Dr Mike Sargent – for being the key inspiration for the visit and for securing the financial support through the ATSE to whom we are also grateful.

Thanks must also go to Dr Ewan Birney who took time out from his busy schedule to visit the Australian bioinformatics community and provide his valuable contribution to the many debates.

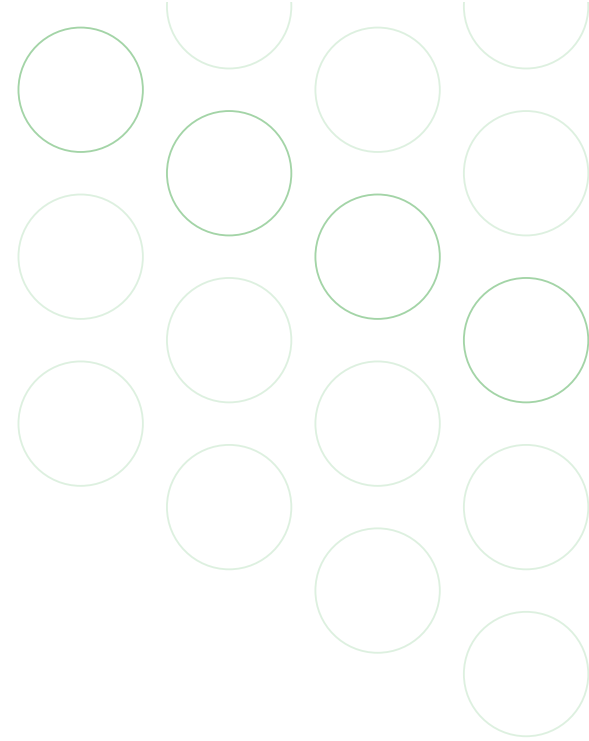
In addition to the engagement and awareness raising aspects of the workshop and seminar series, EMBL Australia hopes that this document will become a useful reference point of current thinking in this area for key decision policy makers and funders.

Silvio Tiziani
Executive Director
EMBL Australia

Summary of Activities by State

In the pages to follow are the details and summaries of Dr Ewan Birney's visits to six states in Australia.

1. Adelaide (South Australia) – compiled by Prof David Adelson
2. Brisbane (Queensland) – compiled by Prof Mark Ragan
3. Canberra (Australian Capital Territory) – compiled by Dr David Lovell
4. Melbourne (Victoria) – compiled by Prof Paul Bonnington
5. Perth (Western Australia) – compiled by Prof Peter Leedman
6. Sydney (New South Wales) – compiled by Dr Jonathan Arthur and Prof Marc Wilkins



Adelaide Bioinformatics for the Future

Biolncubator South Australia
40-46 Thebarton Road, Thebarton, SA

Thursday, 5th March 2009
12.00 noon to 3.00pm

List of Attendees

Abbott, Cathy	Flinders University
Bowen, Joanne	NHMRC Research Fellow, Mucositis Research Laboratory, IMVS
Carroll, Jacqueline	Directorate Manager, Molecular Pathology, IMVS
Coddington, Paul	Deputy Director, eResearch SA
Curtin, Chris	Senior Research Scientist, Australian Wine Research Institute
Gecz, Josef	Professor (Human Genetics), University of Adelaide
Goodall, Greg	Associate Professor Department of Medicine/Head, Cytokine Research Laboratory, IMVS
Hack, Jeremy	Technical Officer Microbial Metabolomics, Australian Wine Research Institute
Li, Jiuyong	Associate Professor, School of Computer and Information Science, UniSA
Mercurio, Meagan	Coordinator, Microbial Metabolomics, The Australian Wine Research Institute
Mitchell, Jim	Head of Biological Sciences, Flinders University
Murrell, Ken	Principal Policy Officer, Science and Innovation Directorate, DFEEST
Roberts, Claire	Associate Professor Research Centre for Reproductive Health, University of Adelaide
Rodgers, Ray	Research Centre for Reproductive Health, University of Adelaide
Rudzki, Zbigniew	Head, Department of Molecular Pathology, IMVS
Solomon, Patty	Chair, Bioinformatics, University of Adelaide
Stanley, Andrew	Policy and Inter-Government Relations Division Department of Health (DoH)
Szubert, Marek	IMVS Bioinformatician
Tizard, James	Director (Interim), eResearch SA/CEO, SABREnet

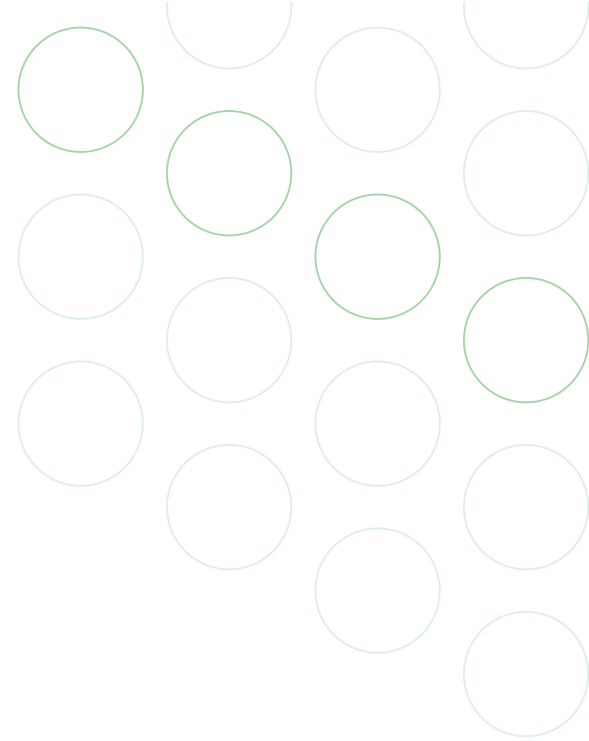
Agenda Items

11.00am – 12.00pm	Seminar	Ewan Birney
12.00 – 12.45pm	Light lunch and introductions	
12.45pm	The Round Table Introductions	
12.50 – 1.45pm	Presentations	Ewan Birney Dave Adelson
1.45 – 3.00pm	Open Discussion	

Key Points

Participants felt that bioinformatics is extremely important to Australia and that there is an extreme shortage of skilled bioinformaticians and computational biologists. They believed this is only going to worsen with the adoption of next-gen DNA sequencing. The unanimous view was that there is currently an analysis bottleneck in biology caused by a lack of bioinformaticians. Discussion focused on how to overcome this shortage of trained personnel.

1. *Best way to increase bioinformatics infrastructure.* Participants debated the merits of altering curricula to increase the number of locally trained computational biologists vs. attracting people from physics and maths. There was a consensus that Biology degrees must include more maths and statistics to train the next generation, but that, in the short term, the right staff need to be hired in order to attract good students and train them.
2. *Centralized vs. distributed infrastructure, i.e. outsourced to core facility with consultants/cost recovery vs. embedded staff funded by grants.* There were both positive and negative aspects to these alternatives presented, with opinions slightly favoring the embedded scenario. Concerns were raised about career paths for embedded bioinformatics personnel and about how they could become isolated if they were not part of a critical mass. It was not clear to people how to reconcile a distributed bioinformatics infrastructure with critical mass.
3. *How to fund additional personnel.* The consensus was that both the Federal government and the states must increase funds to Universities/Research institutions to make new hires. These funds cannot be a one-time injection but must be linked to ongoing, increased research support in the area to provide career paths and encourage students to enter the field.
4. *Cup half full.* While everyone acknowledged that the current situation is dire, the advent of next gen sequencing technology levels the playing field in biology, making it far more likely that increased investment in bioinformatics will be money well spent and promote new discoveries of significant economic impact in health and the resource sector. Because data is freely accessible, it is possible for Australia to become internationally competitive in bioinformatics by attracting existing researchers and training new ones. Funding is the limiting factor on this front.
5. *International collaborations.* These were viewed as essential in the short term in order to surmount the current analysis bottleneck and to provide overseas training.
6. *View of SA DFEEST.* A representative from the State government indicated that they view bioinformatics as a priority area within the research and innovation portfolio and hope that the Federal Government feels likewise.



Brisbane Bioinformatics for the Future

All Ords Auditorium
Queensland Government
102 George Street, Brisbane, Qld

Thursday, 12th March 2009
12.00 noon to 3.00pm

List of Attendees

Barendse, Bill, Dr	Molecular Genetics CSIRO Livestock Industries
Barker, Jeremy	CEO, Queensland Facility for Advanced Bioinformatics
Brown, Margaret	Consultant, Minter Ellison
Bushell, Gillian, Prof	Dean (Research) Science, Environment, Engineering and Technology, Griffith University
Chalk, Alistair, Dr	Research Fellow, Systems Biology Program, Eskitis Institute, Griffith
Clark, Georgina, Dr	Immunoregulation Team Leader, Mater Medical Research Institute Medical Research Institutes
Cochrane, Tom, Prof	Deputy Vice-Chancellor, Queensland University of Technology
Conway, Topaz, Dr	Deputy Director (Development) Mater Medical Research Institute, Medical Research Institutes
Crowe, Mark, Dr	Acting Production Manager, Pfizer Animal Genetics
Cuthbert, Brad	Principal Project Officer Office of Biotechnology, DTRDI, Queensland Government
Dalrymple, Brian, Dr	Science Leader Bioinformatics, CSIRO Livestock Industries
Dinger, Marcel, Dr	Genomics and Computational Biology, The University of Queensland
Edwards, Dave, Dr	Bioinformatics Leader, Australian Centre for Plant Functional Genomics, University of Queensland
Fisk, Nick, Prof	Director, Centre for Clinical Research, The University of Queensland
Gonda, Tom, Prof	Director of Research Diamentina Institute for Cancer, The University of Queensland
Gorse, Dominique	Technical Manager, Queensland Facility for Advanced Bioinformatics
Gray, Jill, Dr	Manager, Technology and Commercialisation, DTRDI, Qld Government
Grieve, Paul, Dr	Manager, Emerging Technologies Queensland Dept of Primary Industries and Fisheries
Grimmond, Sean, A/Prof	Genomics and Computational Biology, The University of Queensland
Hansen, David, Dr	Principal Research Scientist and e-Health Theme Leader, Australian e- Health Research Centre, CSIRO
Hogan, James, A/Prof	School of Software Engineering and Data Communications, and Microsoft e-Research Centre, Queensland University of Technology
Jacobs, Mark, Dr	Director, Office of Biotechnology, DTRDI, Queensland Government
Kaplan, Simon, Prof	Executive Dean, Faculty of Science and Technology, QUT
Kromer, Jens, Dr	Manager Australian Institute for Bioengineering and Nanotechnology, UQld
Mortimer, Robin	Senior Director, Office of Health and Medical Research Queensland Health, Queensland Government
Pailthorpe, Bernard, Prof	CEO, Queensland Cyber Infrastructure Foundation
Ragan, Mark, Prof	Head, Genomics and Computational Biology, The University of Qld Director, ARC Centre of Excellence in Bioinformatics
Raymond, Kerry, Prof	e-Science Program, Faculty of Information Technology, QUT
Reverter-Gomez, Antonio, Dr	Bioinformatics, CSIRO Livestock Industries

Roach, Leigh	Executive Director Technology and Emerging Industries, Queensland Government
Russell, Paul	Executive Director Enabling Technologies, DTRDI, Queensland Government
Scott, Paul, Dr	ARC Centre of Excellence for Integrative Legume Research The University of Queensland
Shannon, Elizabeth	Manager, ICT, Industry Development, DTRDI, Queensland Government
Siddle, David, Prof	Deputy Vice-Chancellor, The University of Queensland
Stowe, Jenny, Prof	Director of Research Institute for Molecular Bioscience, The University of Queensland
Thomas, Mervyn, Dr	CEO, Emphron Informatics
van Niekerk, Alvin, Dr	COO, ARC Centre of Excellence for Integrative Legume Research, The University of Queensland
Wong, Lanna	Manager, ARC Centre of Excellence in Bioinformatics, UQld

Agenda Items

Thursday 12th March		
9.00 – 10.00am	Seminar – IMB Genomics and	Ewan Birney
10.00 – 10.30am	Discussion	
12.00 – 3.00pm	The Round Table Discussion	
12.35 – 1.30pm	Presentations	Prof Nadia Rosenthal, Dr Ewan Birney, Prof Mark Ragan
Friday 13th March		
12.00 – 1.00pm	IMB Seminar – ENCODE, Ensembl	

Key Points

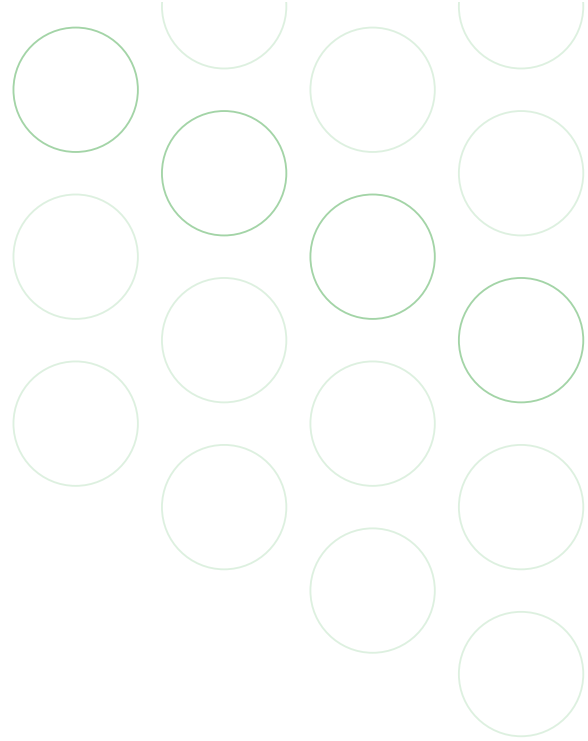
Dr Ewan Birney presented the challenges in bioinformatics, as well as perspectives from EBI and on how to fund and attract people.

Understanding Bioinformatics

- Increase in data-driven “dry” biology since the 1980s
- Technology originally focussed on DNA
- Imaging technology increasingly useful in biology
- Increased volume from new technology for DNA sequencing
- Very large-scale experiments are now feasible
- Continued shift towards “dry” science
- Bioinformatics represents majority of cost in genome projects
- All modern genome projects now have an analysis bottleneck
- Bioinformatics 1990–2010 – driven by computer power

Costing Themes

- Easier to justify and allocate machine costs than salary within granting systems
- Modern bioinformatics budget will be 20:80 hardware:salary (perhaps moving toward 25:75)
- Availability of bioinformatics skills seen as a hurdle
- Hard to find people, hard to build critical mass
- Mid-to-long term, we need to change outlook of “quantitative/maths” students away from default focus on physics and chemistry, adding biology as a 1st tier quantitative science, akin to statistical physics
- Wet/dry linkage can inspire recruitment & increase intake of students



Types of Bioinformatics

- Infrastructure
- Production bioinformatics
- Analysis bioinformatics

Infrastructure

- Often used only within an organisation
- Linkage to public data sets
- Needs specific funding

Production bioinformatics

- Requires large-scale machines (especially for next-generation sequencing)
- Large capital spend
- Generally under-resourced in bioinformatics for next-generation sequencing
- Primarily needs people/skills on the dry (computational) side
- Ideal next-generation sequencing installation would have 50 TB of “scratch” disk space, and computing capacity around 100–200 cores

Analysis bioinformatics

Two models:

- embedded within the research team
- outsourced (“consultant”)

Embedded model

- Raise funds from grants
- Bioinformatics personnel included in grants
- Focus on science problems

Pros

- Raise funds from grants, therefore focus on science problems
- Creative “wet/dry” environment

Cons

- Bioinformaticians can be isolated from critical mass / cognate skills
- Principal Investigator needs to understand bioinformatics enough to direct this facet
- Reviewers of grants need to understand the need for bioinformatics

Outsourced model

- Centralised group from which time is purchased by the researcher

Pros

- Mix of complementary bioinformatics skills and capabilities
- Can develop critical mass
- Often easier to recruit & retain personnel
- More-efficient use of scarce resources

Cons

- Often brought in at the end of a project, not from the start, which limits effectiveness
- Tends not to engage in the science or help define the research question
- Task given – turning the bioinformatics handle
- Funding needs to be underwritten for a good bioinformatics team to be built – core skills

Change by 2010–2020

Genome-wide association studies will be well-established. Although there is no obvious genomics impact on health care currently, it is very likely within 5 to 20 years that there will be. This means that many of the skills (health, IT) in bioinformatics will be transferred into primary health care. It is not just another “IT model” in an IT strategy. Complex systems are likely to be needed. There will be interaction linking genomics research with national strategies on electronic health care records, resulting in an associated capacity problem.

It was intended that problems surrounding bioinformatics capacity would be solved by the NCRIS process. NCRIS is unique to Australia and its funding structure.

Prof Mark Ragan presented on the proposed Australian mirror of EBI, data size in molecular and “omic” bioscience, gap analysis, data quantity, closing (eliminating) the gap, and strategic need for a facility for genome-scale molecular bioscience hardware.

Mirror = enhanced local service centre

What is in it for EMBL/EBI?

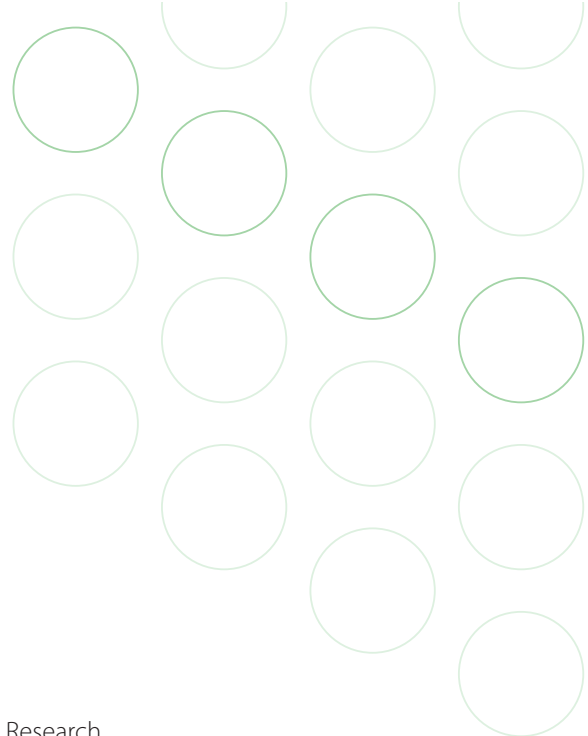
- Distribute demand load, especially from growth in Asia (traffic load, time-zone coverage)
- Engage new communities
- Interface with Australian specialised research such as plant genomics
- Disaster recovery

What is in it for Australia?

- Access to top-class bioinformatics
- Bioinformatics services locally available (avoid data transmission, queue times etc.)
- Competitive edge for local researchers
- Recruitment potential to Australia & institutions
- Opportunity to access advanced training at EBI
- New career development for local researchers
- Leverages large-scale ICT / infrastructure spending in Australia

What and when?

- Research context
- Enhanced data and informatics
- Technology core
- Physical structure
- Petabyte storage



Canberra Bioinformatics for the Future

Industry Link Room
CSIRO Discovery Centre, ACT

Friday 6th March 2009
12.30pm to 3.30pm

List of Attendees

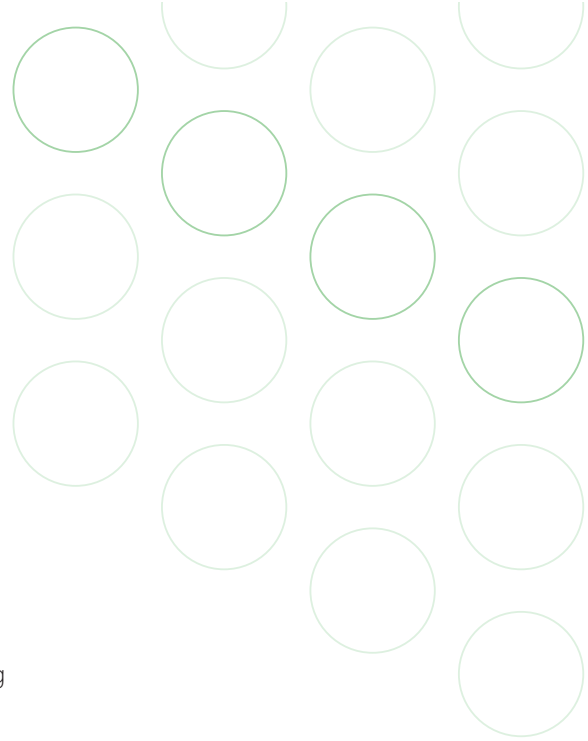
Aitkin, Alex	Department Innovation, Industry, Science and Research
Burden, Conrad	ANU Centre for Bioinformation Science
Collins, Grant	ADFA Chemistry
East, Peter	CSIRO Entomology
Easteal, Simon	ANU JCSMR
Edwards, Owain	CSIRO Entomology
Gordon, Karl	CSIRO Entomology
Graves, Jenny	ANU RSBS
Huttley, Gavin	ANU JCSMR
Jermiin, Lars	University of Sydney
Kilian, Andrzej	Diversity Arrays Technology Pty Ltd
Lovell, David	CSIRO Mathematical and Information Sciences
McNevin, Dennis	University of Canberra – Forensics
Oakeshott, John	CSIRO Entomology
Palmer, Stephanie	ANU Biomolecular Resource Facility
Ryan, Louise	CSIRO Mathematical and Information Sciences
Saint, Rob	ANU RSBS
Saunders, Ian	CSIRO Mathematical and Information Sciences
Taylor, Jen	CSIRO Plant Industry
Tiziani, Silvio	Australian Regenerative Medicine Institute
Wakefield, Matthew	WEHI Bioinformatics
Williams, Rohan	ANU JCSMR
Wilson, Sue	ANU Centre for Bioinformation Science

Agenda Items

Friday 6th March		
11.00am – 12.00pm	Seminar – ENCODE, Ensembl and Short Read Assembly	Ewan Birney (Optus Lecture Theatre, CSIRO Discovery Centre)
12.30 – 3.30pm	The Round Table Discussion –	(Industry Link Room, CSIRO Discovery Centre)
4.00 – 5.00pm	Meet with Margaret Sheil CEO Australian Research Council	1st Floor, 8 Brindabella Circuit Brindabella Business Park <i>Attendees: Silvio Tiziani, David Lovell</i>
5.30 – 6.30pm	Meet with Owain Edwards and Karl Gordon (CSIRO Entomology)	Meeting Room 2, Qantas Club Canberra Airport <i>Attendee: David Lovell</i>

Key Points

1. Biological research is becoming increasingly data-driven (“dry”)
 - Bioinformatics and other data analysis can account for 50-90% of costs
 - Even in the most computationally intensive projects, salary costs dominate
 - All modern genomics projects are now “analysis bottlenecked”
 - Bioinformatics tends to be under-resourced
2. Australia does not yet have a coherent bioinformatics infrastructure
 - EMBL and EMBL Australia’s role in the Australian bioinformatics environment has great potential, but is still evolving
 - Questions remain about how institutions might best engage in EMBL Australia
 - The Australian National Data Service and Australian Research Collaboration Service could play an important role in this, along with other NCRIS initiatives, most notably BioPlatforms Australia
3. There is a global skills shortage in bioinformatics and computational biology, compounded by:
 - Increasing need for “dry” biology in the life sciences
 - A lack of appreciation of this shift towards much more data-driven biology
 - A lack of funding models and career paths that support computational biology
 - The perception that “biology is the science with no maths”
4. Influx and retention of expertise in bioinformatics and computational biology could be encouraged by:
 - Encouraging more students with quantitative skills towards the biosciences
 - Ensuring principal investigators are comfortable with both “wet” and “dry” biology
 - Ensuring that grant proposals are reviewed by people who have an appropriate appreciation and perspective of the domain
5. The last decade has seen a tendency towards short-termism and risk-aversion in the support for Australia’s science
 - The EMBL model of 9-year terms for Group Leaders helps foster scientific excellence
 - ANU used to have research positions of comparable terms
6. Computational biology and bioinformatics are enabling disciplines
 - Methods and data become more valuable when you freely share and combine them
 - The domain changes too quickly and has too small a user base to create a viable software industry
 - Commercial imperatives in computational biology do not necessarily well-serve Australian science
 - Demand is for expertise rather than software
 - We don’t have to choose between commercial benefit and public good; we simply get no benefit if we try to keep methods and data under wraps
7. There are a range of organisational models for bioinformatics and computational biology capabilities, ranging from centralised through to embedded
 - It appears that there is no one “right model”, but a degree of diversity and a flexible approach can help good models to take hold
 - Trust and respect between disciplines is essential for any interdisciplinary activities to thrive
8. Bioinformatics can be thought of as having three flavours
 - Infrastructure – the hardware side of bioinformatics, including sequencers and other measuring instruments, and the IT infrastructure needed to support them
 - Production – activities “close to the machine” that produce digital measurements from biological samples
 - Analysis – activities “close to the science”, involving the analysis of data that have been measured, including their integration with other information
 - Infrastructure and Production bioinformatics need to be closely coordinated



Melbourne Bioinformatics for the Future

Sir George Lush Room
Monash University, Clayton Campus
Tuesday 10th March

List of Attendees

Applebe, Bill, Dr	Victorian Partnership for Advanced Computing
Bacic, Tony, Prof	University of Melbourne
Black, Steve	Victorian e-Research Strategic Initiative
Bonnington, Paul, Prof	Monash e-Research Centre
Borda, Ann, Dr	Victorian e-Research Strategic Initiative
Bowtell, David, Prof	Peter Mac Research
Buckle, Ashley, A/Prof	Monash University
Chen, Phoebe, A/Prof	Deakin University
Chetty, Madhu, Dr	Monash University
Cocks, Ben, Dr	Department of Primary Industries
Coppel, Ross, Prof	Monash University
Forrest, Sue, Dr	Australian Genome Research Facility
Francis, Rhys, Dr	Platforms for Collaboration
Irving, James, Dr	Monash University
Konagurthu, Arun, Dr	University of Melbourne
Kosten, Mark, Dr	La Trobe University
Lefevre, Christophe, Dr	Monash University
McGrath, Annette, Dr	Australian Genome Research Facility
O'Callaghan, John, Prof	Pangalax
Papenfuss, Tony, Dr	Walter and Eliza Hall Institute of Medical Research
Prathjen, Peter, Prof	University of Melbourne
Ragan, Mark, Prof	University of Queensland
Sargent, Mike	MA Sargent and Associates
Savin, Keith, Dr	Department of Primary Industries
Sawbridge, Tim, Dr	Department of Primary Industries
Smith, Ian, Prof	Monash University
Smyth, Gordon, Dr	Walter and Eliza Hall Institute of Medical Research
Speed, Terry, Prof	Walter and Eliza Hall Institute of Medical Research
Sterling, Leon, Prof	University of Melbourne
Taylor, Geoff, Prof	University of Melbourne
Tiziani, Silvio	Australian Regenerative Medicine Institute
Viney, Geoff	Department of Primary Industries
Wakefield, Matthew, Dr	Walter and Eliza Hall Institute of Medical Research
Wallace, Iain, Prof	Victorian Partnership for Advanced Computing
Webb, Geoff, Prof	Monash University
Whisstock, James, Prof	Monash University
Wilkinson, Ross, Dr	Australian National Data Service
Zobel, Justin, Prof	University of Melbourne, National ICT Australia

Agenda Items

Tuesday 10th March		
10.00am – 12.00pm	Public Seminar <ul style="list-style-type: none">• Ensembl and ENCODE: Understanding our Genome• The Australian EBI Mirror Initiative• M Bio Precinct• The e-Research Landscape: Local and Global	Ewan Birney (Lecture Theatre SG01, Bldg 11, Monash University)
12.00pm	Light Lunch	The Sir George Lush Meeting Room Ground Floor, Building 3A, Monash University
1.00pm – 3.00pm	The Round Table Discussion	The Sir George Lush Meeting Room Ground Floor, Building 3A, Monash University

Key Points

Funding for Bioinformatics in Australia

- Australian science funding processes need to recognise the contribution that bioinformatics and dry science is making to life sciences research worldwide.
- This means providing
 - necessary infrastructure support, which could be achieved through NCRIS and an EBI Australia working together,
 - funding for production and analytical bioinformatics as a large component of life sciences grant proposals
 - large grants for large projects that can make significant impact
- First class computer science research applying data mining and other techniques to the massive datasets produced by biomedical equipment.
- Joint work with EMBL/EBI providing access to Australian Government funds provided for involvement in EC FP7 projects.
- Collaboration between leading institutes to generate joint funding for large focussed bioinformatics projects, facilitated by an EBI Australia.

Careers and Training for Bioinformatics

- Australian life sciences will suffer from the global scarcity of trained bioinformatics professionals and computational biologists.
- Establish the perception that computational life sciences is one of the most challenging fields for the application of mathematics and information technology with the aim of attracting a much larger intake of maths-oriented student.
- Build educational infrastructure to identify and cross-train promising students in computational science and especially life-sciences.
- Career paths for bioinformatics professionals – the 80% of bioinformatics practitioners who provide technical assistance to scientists – need to be developed to make the field attractive.
- Develop Principal Investigators who have wet/dry experience.

Establishing a Bioinformatics Structure in Australia

- There is a move from small embedded groups of bioinformatics experts in life science research teams to larger outsourced groups.



- This is viewed as helpful in attracting talent and in making computational biology and production bioinformatics available to a growing group of life science researchers.
- It also encourages the potential for establishing a strong collaboration between the emerging outsourced groups, perhaps around an EBI Australia.
- And forms the basis of a critical mass for building unique components of bioinformatics infrastructure, such as mirrored global datasets and unique Australian contributions to global dataset development.
- This represents an organic growth model which can be encouraged through NCRIS and looks most likely to succeed in Australia.

Strategy for Interaction with EMBL/EBI

- Australian bioinformatics needs to develop its value proposition to EMBL/EBI to facilitate the rapid introduction of Australia into European projects.
- This could include the availability of infrastructure such as marsupial and plat biology datasets, capability such as data mining and interests in areas of focus.
- Identify bioinformatics projects in FP7 where Australian science can make a contribution and work with EMBL/EBI to secure entry – then fund via Australian Government scheme.
- Joint conferences – investigate funding for travel to establish new initiatives.
- Focus Australian bioinformatics on a small number of significant and impactful areas rather than diffusing effort over a large number of small ones.

Introductory Remarks

Mike Sargent

Mike introduced the roundtable as an opportunity to gather information from Australia's leading bioinformatics practitioners. This will contribute to a paper to the Australian Academy of Technological Sciences and Engineering (AATSE), which will also be provided to Margaret Shiel at the ARC.

Following the formation of EMBL Australia with contributions from six institutions and the NCRIS program, Australia is formally linked into EMBL in Europe. This presents great opportunities including working collaboratively with EBI, developing strong relationships with EMBL and EBI and giving something back to EMBL.

Ewan Birney

This preface is based on experience working with a number of bioinformatics developments in Europe, particularly in the UK, Paris and Norway. Everyone is heavily involved in the transition to “dry” computational biology and no-one has got it right yet. Some of the issues are:

- Grants in genomics request an average of 50% and sometimes up to 80-90% of the money for bioinformatics, and similar trends are evident in other areas of bioscience. EBI spends 20% on capital expenses and 80% on operational expenses. This may trend towards 25:75.
- Increasing use of advanced biomedical equipment is generating vast amounts of data at an increasingly rapid rate. This is most evident with DNA sequencers and imaging.
- All modern genomics projects are analysis bottlenecked – plenty of examples.
- Projects are continuing to embrace computational power, from one-person-and-a-desktop years ago, to the immediate future where some projects will have 200 FTEs and petascale storage and computing, and most will have 10 FTE informatics personnel and terabyte scale.
- Infrastructure for bioinformatics includes computation and the large public bioscience datasets are being created and mirrored for researchers around the world. These are so large that there is likely to be at most one in Australia. Enormous archival data repositories are being created.
- Production bioinformatics delivers bioinformatics capability to bioscientists. These people are usually sited close to the infrastructure and are invaluable in delivering results. Grants need to fund salaries first and machinery second. A typical research set-up might include 50Tb of scratch disk, 100–200 cores, 2 production bioinformaticians and one systems person.
- Analytical bioinformatics is where the science happens, results get interpreted and the debates are joined. This is the preserve of computational biologists.
- There is a great scarcity of good people worldwide. This has a higher impact in Australia because of the smaller scale. You can wait to recruit new people when you already have 300; you can't when you only have three.

- Mathematics is becoming vital to biology – in fact biology is becoming one of the most exciting places to do maths.
- Principal Investigators who have a deep understanding of both wet and dry biology are needed to inspire the development of productive bioinformatics.
- Australian bioinformatics has to consider the models that it uses to provide bioinformatics to sciences – the embedded and outsourced models
- In the embedded model, bioinformatics FTEs are generally obtained as a result of grant funding.
Pros: increased focus on the science and balance between wet and dry science in the projects.
Cons: bioinformatics people are isolated from their peers and from tools they might be able to access, PIs are often not sufficiently bioinformatics-aware and grant reviewers do not necessarily understand bioinformatics and are not prepared to allocate sufficient funding.
- In the outsourced mode, a central group is funded separately often on a user-pays basis. Grant applications need to include provision for using outsourced bioinformatics.
Pros: a critical mass of people and better candidates for the positions in a larger organisation; efficiencies in the use of tools and resources
Cons: outsourced groups are often brought in too late in a project, are not sufficiently engaged in the science, are underfunded and are not part of the team.
- EBI is primarily infrastructure with people who participate in the analysis. There is zero production bioinformatics with no wet labs. By contrast the Sanger Institute across the road is production only. The EBI model is neither embedded nor outsourced, it is pure bioinformatics.

In concluding, Ewan pointed out that the impact of bioinformatics on health care over the period 2010 to 2020 is likely to be considerable. It will have a huge impact on clinical research and this will translate to the doctors' surgery and personalised medicine.

Discussion – Funding for Bioinformatics Research in Australia

Issues that were voiced included:

- In the UK the Medical Research Council (MRC), the Biological and Biotechnology Research Council (BBSRC) and the Wellcome Trust provide funding for bioinformatics. Proposals are evaluated by a composite of biomedical and informatics peer review panels including people with bioinformatics expertise.
- This contrasts with the ARC (and probably with the NHMRC) where bioinformatics proposals tend to get batted backwards and forwards between bio and informatics panels and often fail to get funded even though the quality was high.
- Proposals involving large bioinformatics datasets need budgets of several million AUD, making it unlikely that the ARC and NHMRC will fund them.
- As a consequence of these factors bioinformatics work is often hidden in proposals for life sciences research. Bioinformatics needs to be made explicit and to become expected as a core component of life sciences proposals.
- It is hard to create proposals which include both first class bioscience and first class computer science. One is usually the servant of the other.
- Nevertheless bioinformatics is very fruitful prospecting territory for interesting computer science problems. Examples are the cleansing of large amounts of noisy, "dirty" data, and the discovery of unsuspected patterns by mining large datasets and combinations of datasets.

This discussion suggests that researchers need to construct bioinformatics proposals carefully for submission to biosciences and informatics panels, and that panels need to be better informed about the importance of bioinformatics with more bioinformaticians should sitting on panels.

Ewan suggested that if grant sizes are capped, the decreasing costs of sequencing and other biomedical equipment and procedures should provide some headroom for increasing funds requested for bioinformatics apex in life sciences proposals. A bold move would be to seek funding for consumables plus bioinformatics expertise only.



NCRIS 5.1 had provided funding for bioinformatics infrastructure including people delivering technical assistance. Although the results had not won universal acclaim future NCRIS rounds provide an alternative and complementary source of funding for bioinformatics from the ARC/NHMRC research method.

Europe has a pattern of funding large institutes like EMBL and EBI that then work with universities to establish concentrations of research and to cultivate the careers of leading researchers by providing paths from universities to institutes and back.

Australia, in contrast, funds groups at universities and expects university groups to form collaborations which might achieve the same aims as the European institutes. This structural difference is unlikely to change rapidly, which makes it interesting to consider how collaboration with EMBL and EBI could seed the concentration of research in Australia and the cultivation of careers.

Careers, Structure and Training for Bioinformatics

There is debate about whether the best bioinformaticians come from a biosciences or a computer science background. Is it easier to train bioscientists, who often lack strong mathematics, in the analytical side of bioinformatics, or to take someone with a strong computing background and inculcate bioscience intuition and practice? Suggestions were to cross-train students in both disciplines, and to build education into bioinformatics infrastructure to generate a broader corps of expertise amongst young scientists.

The image of bioscientists needs changing so that bioscience attracts students with a strong mathematical background. "Dry" bioscience is one of the most challenging fields today for mathematically talented people. Traditionally those who are scared of maths do bioscience.

A key requirement for bioinformatics success is creating a cadre of Principal Investigators with the right backgrounds, skill-sets and management capabilities. This involves taking risks to inject excellent people with the right leadership potential who come from IT, physics and maths into biosciences.

Training is different for bioinformatics research – call these people computational biologists, and for bioinformatics infrastructure – bioinformatics professionals. The need for bioinformatics professionals is strong enough to establish an education track and a career track for future experts. Who provides the funding for this type of training?

Experience shows that pure-research, theoretical computational biologists make up perhaps 10–20% of the bioinformaticians required and the remaining 80-90% are professionals working in the context of a wet laboratory project.

Bioinformatics people will develop and use techniques for dealing with extremely large volumes of data and for creating, adapting and running analytical models. Experimental "wet" science will be conducted in parallel with computational "dry" science.

Establishing a Bioinformatics Structure in Australia

Embedded bioinformatics people who have access to large computational resources are being superseded in Australia by outsourced groups like VLSCI which provide an expert team of computational biologists and bioinformatics professionals who can work for long periods of time with life science teams to build and run large analytical models over very large amounts of data.

This model has parallels at the Queensland Facility for Advanced Bioinformatics where 14 staff provide a service to a wide range of researchers from a number of institutions on a user pays basis. The Victorian Bioinformatics Consortium works in a similar fashion. There was a feeling that the model suits production bioinformatics well.

The debate centred around the creation of critical masses of bioinformatics expertise, both computational biologists and bioinformatics professionals, and how to create and operate them effectively. In Europe EBI has developed a model for creating a critical mass (of about 350 people) which is large enough to be self-sustaining and linking its expertise to the best universities and research institutions in Europe and beyond. In Australia groups are limited to 30 people or less, are sub-critical size and struggle to gain further momentum by collaborating.

Could a critical mass be created in Melbourne or Parkville around VLSCI? We probably need around 100 bioinformatics people to form a critical mass, and maybe 10 groups of this size around the country. WEHI already has 30 bioinformatics people and there may be as many as 60 or 70 in the Parkville Precinct.

A collaborative approach growing the resources available in Australian centres and then linking them into a distributed national group seems practical, and a continuation of trends in progress today. This can be catalysed by the collaboration with EBI and EMBL, providing opportunities for mutual exchanges of datasets, people and bioinformatics knowledge and know-how. This would be an example of an effective organic growth model.

Strategy for Interaction with EMBL/EBI

EBI would like to see a collaborative bioinformatics drive in Australia to create unique infrastructure datasets for general global use. Examples would be marsupial datasets, plant biology for Australian plants and epidemiological and public health data sets where Australia has some potential for leadership. Health records data linkage may be helpful.

From an Australian point of view this means finding a fit with EMBL through an Australian unique selling proposition that provides value to EMBL. This can include the unique Australian datasets listed above, unique capabilities available in Australia, for example data mining expertise with deep experience in bioscience data or opportunities to work on animals or plants that are not common in Europe – sugarcane for example.

There are sources of funding for joint work between Europe and Australia which can be facilitated by EMBL at both ends. The European Framework 7 projects are being launched using European Community funds for European participants. Australian organisations can now join Framework 7 projects, using funding established by the Australian Government (DIISR) to finance Australian participation.

Joint conferences will also help to seed understanding, collaboration and interactions.

It is important for Australia to understand its capacity and capabilities. It would be preferable to concentrate on a smaller number of leading edge projects concentrating on major areas of interest, such as major plants.

Facilitator's Summary

Bioinformatics is a key part of life sciences research and represents a major component of the cost of most projects. Dry research is steadily increasing relative to wet, and is becoming more pervasive across agriculture and health as well as medicine.

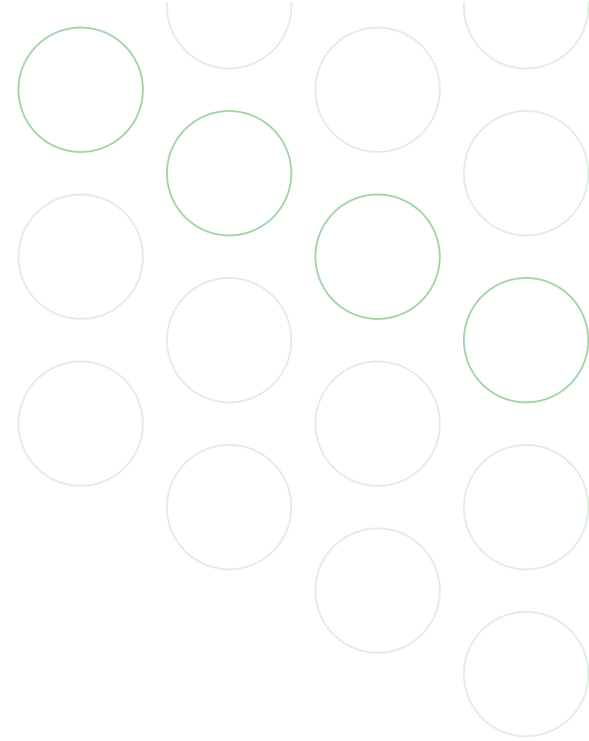
Australia will inevitably become a player on a petabyte scale.

Skills and career structure for bioinformatics are a significant issue. Australia needs to evolve a structure that achieves critical mass while retaining strong identification with the individual research drivers.

For research excellence there needs to be a focus on the analytical component of bioinformatics, and on developing excellence in production bioinformatics in some areas, for example population health studies. Bioinformatics should seek to benefit from the existence of ANDS and present examples of petascale data management issues for ANDS attention.

An outreach program is important. It is important that research funding agencies and reviewers understand the importance of bioinformatics to the life sciences, and that some consideration is given to increasing the sizes of the projects that they fund.

Finally, bioinformatics in Australia needs to work towards a national collaborative arrangement with strong international links especially through EMBL. A structure for bioinformatics interfaces with the international world such as an EBI Australia is key.



Perth Bioinformatics for the Future

The Western Australia Institute for Medical Research
Queen Elizabeth II Medical Centre Campus
Level 5 and 6, MRF Building
Rear 50 Murray Street, Perth WA

Friday 20th March 2009

List of Attendees

Bellgard, Matthew, Prof
Carter, Kim, Dr
Datta, Amitava, Prof

De Kerk, Nick, Prof
Giles, Keith, Dr
Laing, Nigel, Prof
Leedman, Peter

McEachern, Doug, Prof
Millar, Harvey, Prof
Ravine, David, Prof
Stefanov, Valery, Prof
Tonti-Filippini, Julian

Centre for Comparative Genomics, Murdoch University
The Telethon Institute for Child Health Research
University of Western Australia, School of Computer Science and Software Engineering (CSSE)
Telethon Institute for Child Health Research
University of Western Australia, Centre for Medical Research, WAIMR
University of Western Australia, Centre for Medical Research, WAIMR Chair, The Western Australia Institute for Medical Research
University of Western Australia, Centre for Medical Research
University of Western Australia, DVCRI
ARC Centre of Excellence, Plant Energy Biology
University of Western Australia, Centre for Medical Research, WAIMR
University of Western Australia, School of Mathematics and Statistics
ARC Centre of Excellence, Plant Energy Biology

Agenda Items

Friday 20th March		
12.00pm – 1.00pm	Public Seminar <ul style="list-style-type: none"> Bioinformatics, EMBL and Australian Researchers 	Ewan Birney (WAIMR Seminar Room, WAIMR Nedlands Campus, Ground Floor, B Block)
1.00pm – 3.00pm	The Round Table Discussion	(WAIMR Seminar Room, WAIMR Nedlands Campus, QEII Medical Centre, B Block)

Key Points

Some of the key take-home messages from the round table included:

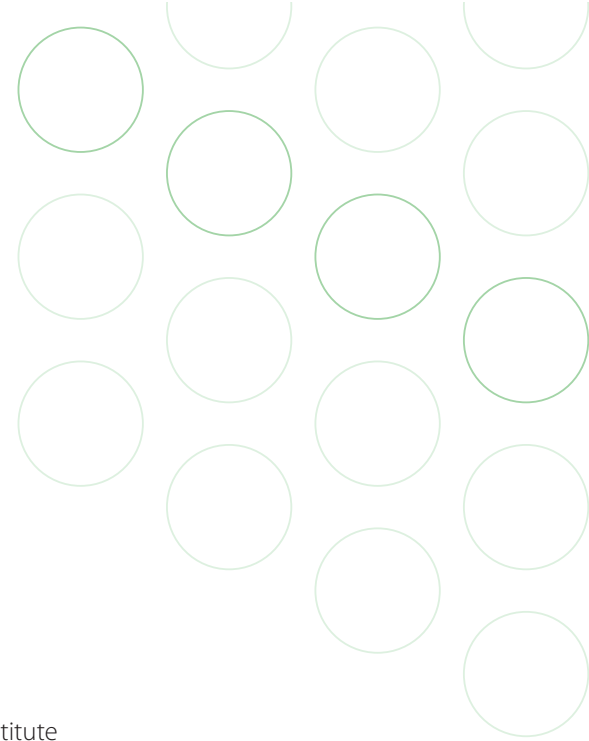
1. Australia is developing a competent bioinformatics sector but needs substantial investment in people in the next five years to have any chance of being truly internationally competitive in the next 20 years.
2. EMBL, the EBI and Ewan Birney are keen to be involved with the growth of Australian bioinformatics, and could substantially “value-add” to Australia’s investment.
3. Institutes and Universities need to lead this growth in informatics by appointing significantly more dry “Bioinformatics” scientists.
4. These are high priority issues that have to be acted upon even in the current global financial crisis.
5. Western Australia with its remarkable population and clinical disease databases presents some very real opportunities for bioinformatics collaborations with EMBL/EBI and internationally.

Bioinformatics for the Future Lecture – WAIMR

Ewan's lecture at WAIMR was the last of his sponsored around Australia EMBL speaking tour, an EMBL Australia initiative and supported by the Australian Academy of Technological Sciences and Engineering and the Australia Department of Innovation and Australian Government Department of Innovation, Industry, Science and Research. It was very well attended and there were multiple questions throughout his talk prompting very fertile discussions.

Round-Table Workshop and Discussion – WAIMR

- Ewan presented a short perspective from EMBL/EBI raising several key themes on informatics that he had outlined in a white paper distributed prior to his arrival.
- A key outcome from this discussion was the realisation that bioinformatics, just like PCR, is becoming an almost "normal" component of most biomedical research. The massive increase in data generation in many biomedical projects requires increasing sophistication in data handling and presents a bottleneck for many laboratories that are not well staffed with bioinformaticians. To address this necessitates considerable up-skilling of biomedical researchers across the board, provision of appropriate people infrastructure to teach these skills widely, and employment of more experienced high-level informatics experts to oversee this whole process.
- The key driver in Australia achieving these goals relates to finding personnel with expertise and not the computer hardware. The EBI experience suggests that about 80% of their costs are for staff and about 20% for hardware and computers.
- Ewan supported the concept that analysis bioinformatics should be progressed in an embedded model which is likely to be the only real long term solution. However, a component of outsourcing to larger well-established informatics groups with a significant critical mass would also be important. Some of these positions could be joint appointments where the Private Investigator may cover multiple different groups and provide bioinformatics in that context.
- Ideally we need to embed new people into long term bioinformatics positions in institutes and academic organizations around Australia, at the same time as we build capacity to have some select groups generate a major critical mass as high quality outsource providers of bioinformatics.
- There was significant discussion supporting the case for an EBI mirror in Queensland, which would provide Australian institutions and researchers with high quality informatics at high speed.
- Although next generation sequencing presents extraordinary possibilities, Ewan emphasised the need for each machine to be appropriately supported with staff with experience in informatics and data processing.
- Ewan emphasised the importance of integrating information technologies and informatics into improving health outcomes.
- There was much discussion about how EMBL Australia and partnership labs might work with EMBL and EBI. Ewan clarified the sorts of opportunities that exist at EMBL/EBI for Australian scientists (visits/workshops/faculty positions/PhD studentships via Australia's Associate EMBL member status etc). Visits to Australia by other EMBL/EBI informatics experts may well also be possible, which would continue to open up new avenues of communication and collaboration.
- There was considerable focus on the population databases in Western Australia. The potential for further exploring these exceptional long-term community databases was discussed with Ewan and EBI. Ewan thought there were some very interesting possibilities, in several areas for collaborative joint EMBL/EBI projects in this area.



Sydney Bioinformatics for the Future

The Medical Foundation Building
University of Sydney
NSW 2006

Wednesday 11th March 2009

List of Attendees

Rosenthal, Nadia, Prof	Director, Australian Regenerative Medicine Institute Director, European Molecular Biological Laboratories
Arthur, Jonathan, Dr	CEO/Director Sydney Bioinformatics, University of Sydney
Wilkins, Marc, Prof	Director, NSW Systems Biology Initiative, University of New South Wales
Dawes, Ian, Prof	Director, Ramaciotti Centre for Gene Function Analysis, University of New South Wales
Wade, Clare, Prof	Professor of Animal Genetics and Computational Biology, University of Sydney
Belov, Kathy, Dr	Australasian Wildlife Genomics Group, University of Sydney
Yu, Bing, Dr	Acting Director, SUPAMAC, University of Sydney
Thomas, Torsten, Dr	School of Biotechnology and Biomolecular Science, University of New South Wales
Sloggett, Clare, Dr	Intersect Pty Ltd
Gilbert, Andrew	Bioplatforms Australia

Agenda Items

Tuesday 10th March		
10.00am – 12.00pm	Public Seminar <ul style="list-style-type: none"> • ENCODE, Ensembl, and Short Read Assembly • Presentation on EMBL Australia • Presentation on NSW Genomics Capacity 	(Medical Foundation Building Auditorium, University of Sydney) Dr Ewan Birney Prof Nadia Rosenthal Prof Ian Dawes
1.00pm – 4.00pm	The Round Table Discussion	

Key Points

The major topic of discussion was the intersection between bioinformatics and high-throughput sequencing. The following major areas were covered:

1. A perspective from the EBI

Dr Ewan Birney provided a second, short presentation developing several themes he had outlined in a white paper distributed to participants in the round table discussion prior to the meeting. In particular, these themes considered bioinformatics for next generation sequencing.

One key theme was that *bioinformatics costs now form the majority of all costs in some research projects*, particularly those involving next-generation sequencing where bioinformatics costs are typically greater than 50% of the total project cost (and sometimes up to 90%). All modern genomics projects are now bottlenecked at the stage of data analysis rather than data production.

Furthermore, the major component of these bioinformatics costs is *salaries for bioinformatics staff, not computer hardware*. The experience of the EBI is that the split in bioinformatics costs between computer hardware and bioinformatics staff is roughly 20% to 80%.

The issue of bioinformatics infrastructure in Australia was discussed. Three broad aspects of bioinformatics infrastructure were identified: public data repositories, production bioinformatics, and analysis bioinformatics. It was agreed probably only one group in Australia should be involved in coordinating the provision of public data sets (for example, mirrors of GenBank, Swiss-Prot, or the establishment and maintenance of similar, specialist resources on behalf of the global research community). Greater challenges and demand exist for *Production Bioinformatics and Analysis Bioinformatics*.

Production bioinformatics

This is the bioinformatics closely coupled with analytical instrumentation. It is relatively generic and pervasive. It is required to appropriately store data generated by instrumentation then convert it from a raw format to a final data type ready for downstream analysis. In most cases, this work can be pipelined. Investment in this area is required to prevent wasting capital equipment and consumables. Most sequencing facilities or services should have this capacity. Ewan noted that it is common to under-resource bioinformatics for next generation sequencing machines. He commented that a small sequencing facility (with two or three sequencing machines) would typically need 100 to 200 cores of processing power combined with 50 TB of disc scratch space. Additionally, 2 full time bioinformaticians and 1 full time systems administrator are required.

Analysis bioinformatics

This bioinformatics comprises the actual biological analysis of data generated by instruments. It is closest to, and most influenced by, the scientific questions being explored. It is where the final analytical data is analyzed. For applications involving next generation sequencing, this includes contig assembly, matching to reference sequences, or the analysis of copies of sequences for transcriptomics. There are two possible models for completing this bioinformatics. Both are currently used in Australia and each has advantages and disadvantages.

Embedded in research teams.

In this model, bioinformatics researchers form part of a fully integrated research group, where “wet” (laboratory based) and “dry” (computer based) biological analyzes are completed side-by-side towards a common research goal. The bioinformatics researchers are funded through research grants in exactly the same way as any other research staff in the laboratory. The advantage of this model is that the bioinformatics staff should be focused on the science at hand and will be able to help manage the trade-offs between wet and dry lab work. They are deeply engaged in the particular research questions being addressed, making it easier to identify novel areas for bioinformatic analysis or undertake targeted development of new algorithms to meet research needs.

The disadvantages are that these bioinformatics staff may become isolated, particularly if they are the only bioinformatics researcher in a large group. This may lead to the bioinformatics researcher failing to leverage existing databases or tools and a lack of critical mass to generate new ideas and approaches. A further issue is access to appropriate computer hardware to undertake sophisticated bioinformatics analysis, which may not be affordable by individual research groups. A final challenge is that, in Australia, and many places in the world, there are relatively few groups that are large enough to have an ‘in-house’ bioinformatician. Furthermore, there is an issue with capacity in that bioinformaticians are difficult to find and some Chief Investigators are not confident to supervise bioinformatics staff.

Outsourcing to “consultants”.

In this model, there is a centralized group of bioinformatics researchers from which consultant time is purchased or otherwise accessed. Sydney Bioinformatics and the NSW Systems Biology Initiative currently pursue this model. This model may attract better quality bioinformatics staff, build critical mass, and provide a more efficient use of resources. However, it is often the case that wet lab researchers will not approach the bioinformatics staff until the end of a project, when data has already been acquired. This can lead to



experiments being poorly designed for the bioinformatics analyses desired, with the consequence of reduced value being obtained from the investment in producing the data.

Other potential disadvantages are that the staff may not fully engage in the science of the collaborators, and may 'turn the handle' for what might become relatively routine analyses. Bioinformatics staff may also find it hard to build independent research interests in this type of setting. This model requires a level of underwriting of funding to be successful.

While both models of bioinformatics have advantages and disadvantages, there appeared to be a general consensus towards embedding bioinformatics in research teams and building fully integrated research groups. It was noted the consultant model may, however, be a good interim solution while integrated groups are built, or an auxiliary solution for small research groups.

2. EMBL Australia and partnerships with EMBL/EBI

The question of how NSW-based researchers could work in association with the EMBL to develop effective bioinformatics for its next generation sequencing installations was also discussed, including:

- Could local staff visit EMBL and EBI to learn from their data management experience?
- Could staff from EMBL and EBI visit the local facilities to share their experience?
- Could any tools, databases, and pipelines developed by EMBL or EBI be transferred to local facilities (technology transfer)?
- Could any hardware environment established by the EBI for this type of analysis be replicated locally?
- Could EBI staff provide training for key aspects of 'analysis bioinformatics' to a local audience?

Major outcomes of this section of the discussion were:

a. There are opportunities for collaboration

EMBL cannot send funds to Australia, as the Australian funds goes into EMBL central funding as per the governance model of EMBL. However, EMBL can send staff to Australia for these types of activities. This is a means of facilitating short visits from key EMBL staff.

There is the 'geek for a week' program in which a scientific collaboration is set up and staff can visit the EBI. This is a project-based collaboration. Travel costs to the EBI need to be funded by the researcher.

b. There is technology and expertise which can be transferred

There is a small sequencing lab in EMBL, separate from the larger Sanger Institute. This small lab runs ~3 next-generation sequencing machines. This setup is likely to be similar in scale to most Australian laboratories or facilities. The EBI has established the production bioinformatics for this type of facility. There is a key individual, Vladimir, who could serve as a key contact in this area. It may be possible to have him visit Australia; equally NSW staff would be able to visit the EMBL to view and evaluate this facility.

c. Local staff could be trained in key aspects of bioinformatics for next-generation sequencing machines.

NSW-based staff could visit the EMBL for this training; however remote training is also a possibility. The latter may be more economical if a number of local researchers could be trained simultaneously. During a visit to EMBL, key technologies could be learned in a relatively short period of time (2 to 3 weeks). Areas of most relevance to Australian researchers probably differ from those of large sequencing projects. Training in the following areas would be more relevant:

- RNAseq/ChIPseq work
- Targeted resequencing work
- Metagenomics work
- Variants analysis work

3. Action items arising

- Dr. Ewan Birney and other EBI staff to provide (email) guidance to NSW-based staff on the implementation of production bioinformatics for next generation sequencing
- Prof. Nadia Rosenthal to explore the opportunity for Vladimir to visit NSW to assist in the implementation of production bioinformatics for next generation sequencing
- Dr. Clare Sloggett (Intersect and UNSW) to visit the EBI in April as part of an upcoming trip
- That training into analysis bioinformatics for next generation sequencing be revisited after the establishment of next generation sequencing in NSW.
- That NSW-based genomics researchers be kept informed of progress of the NSW EBI interactions.

Discussion Papers by Dr Ewan Birney

Investment in bioinformatics in Australia: an opportunity

This report was prepared by Dr Ewan Birney, Head of Nucleotide Data, European Bioinformatics Institute, Cambridge UK.

Summary

- Biological sciences have been revolutionised by technologies that generate large-scale datasets
- Australia did not participate fully in the first wave of this revolution
- Decreases in technology costs and increases in national collaborative research infrastructure have created a new opportunity for Australia to benefit from and contribute to world-leading bioscience to a much greater extent
 - However, computational analysis has now become the major bottleneck for progress, both in Australia and worldwide, mainly due to a lack of appropriately skilled personnel
 - Relative to Australia's infrastructure spending, a small investment to implement critical bioinformatics capacity stands to yield a disproportionately large return
- For Australia to seize this opportunity for catalytic change in the biological sciences, I recommend
 - Joint local + national investment in bioinformatics infrastructure
 - National coordination of bioinformatics infrastructure
- I estimate the cost of this investment over 5 years would be approximately \$30M from state governments plus \$30M from the Federal Government.

New technology, new opportunities, new challenges

Biological sciences have undergone a revolution in the last decade, moving from single laboratory, experimentally-driven research with customised apparatus in each lab, to more coordinated activity in which centralised facilities use automated equipment to generate large-scale datasets that support investigations in many laboratories and institutions. The sequencing of the human genome by an international consortium in 2001 is an outstanding example of such large-scale collaborative bioscience.

"Next-generation" DNA sequencing machines have accelerated this revolution. Two years ago, the cost of acquiring a human genome sequence was \$50 million. Today it is under \$50,000 – and still falling fast. New technologies continue to empower advances in both medicine and basic biology.

Australia did not participate fully in the first wave of genomics because the overall cost was very high, and also because it required extreme concentration of funds into a few large centres—a challenge even in large economies like the US. The current landscape is very different: the cost of next-generation sequencing technology is far lower (and thus more-widely accessible); and Australian scientific infrastructure is now more organised (through programs such as NCRIS) resulting in better access to new technologies earlier in their lifecycle. Already the extensive and broad talent base of Australian science is using next-generation DNA sequencing for exciting discoveries.

However, *computational* analysis has now become the major bottleneck for progress, both in Australia and worldwide. This bottleneck is mainly due to a lack of appropriately skilled personnel.

Biological data manipulation is described as *bioinformatics*, *computational biology*, or sometimes *dry biology*, with the common theme being the computational analysis of large, shared datasets. Similar themes are seen in other sciences including physics and chemistry, but the shift towards dry biology over the last decade has been the most extreme. Just as physics became more computationally and theoretically oriented in the first half of the 20th Century, biology in these early years of the 21st Century is changing its emphasis.

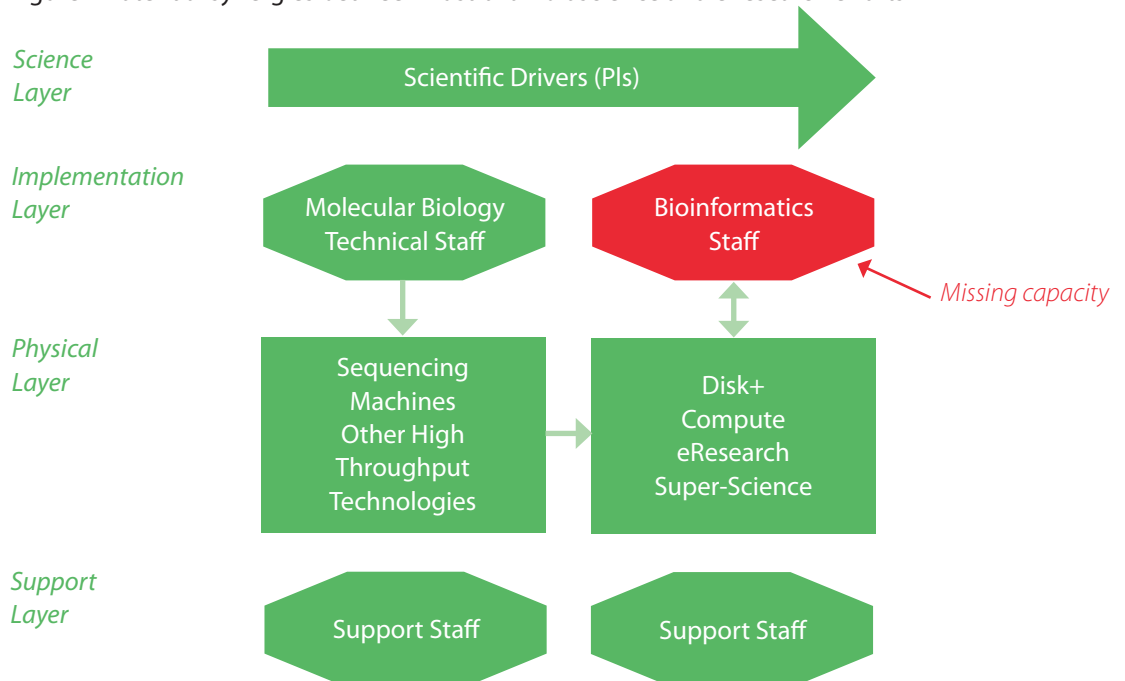
Australia's bioinformatics bottleneck

Catalysed by EMBL Australia, a series of round-table discussions focused on this revolution in biology was held in each state capital and in Canberra in early 2009. Leading researchers from biological sciences, medicine, computation and mathematics, state and national science administrators, and international experts participated. An extensive report with detailed proposals has been prepared by EMBL Australia.

The common theme arising from all the discussions is the incredible demand for more capacity in bioinformatics, and that timely investment in bioinformatics infrastructure—mainly at the level of skilled personnel—would provide a competitive advantage to Australian science, unlock many new discoveries, and increase Australia's presence in the international science landscape.

Strategic investment in skilled personnel in bioinformatics infrastructure would catalyse the potential synergies between the very effective eResearch computational infrastructure (now in transition towards the SuperScience agenda), including national high-performance computing, and these new, far cheaper, next-generation molecular biology technologies (Figure 1). The coordination of high-performance computing in Australia, with a balance between computation and disk provision, is impressive and world-class. Similarly the inventiveness of Australian bioscience has grown and now provides world-leading innovation in a number of areas. *However, the lack of skilled bioinformatics individuals remains the critical missing component in bringing these two world-class programs together.*

Figure 1 Potential synergies between Australian bioscience and eResearch efforts



Relative to Australia's infrastructure spending, a small investment to implement this critical bioinformatics capacity stands to yield a disproportionately large return. It is likely that over five to ten years, this bioinformatics capacity will simply become part of modern molecular biology; at this point in time, however, there is a critical capacity gap and also considerable worldwide competition for skilled individuals. Strategic investment is the only way to enable this science.

To illustrate this in concrete terms, I give three "case studies" of leading Australian science, and their potential to be transformed by investment in bioinformatics.

Case Study 1: Cancer genomics

The Commonwealth of Australia has committed to the intensive study of some 400 pancreatic cancers, and with this commitment has become a major player in the International Cancer Genome Consortium. A variety of different changes are being examined, from somatic (cancer-specific) mutations to amplifications of parts of the genome. The predicted size of this dataset—around 2 petabytes of raw data and around 200 terabytes after screening and compression—is challenging just to capture, store and manipulate, let alone analyse. With increased investment into bioinformatics, sophisticated methods can be developed in Australia to analyse data from this and other very large cancer genome datasets being made available by the International Cancer Genome Project. It is certain that the most effective groups in this area will be the ones with best curated cancer samples (a strength of Australian oncology) and the best bioinformatic analysis.

Case Study 2: Agricultural improvements in plant genomics

New high-throughput technologies have now opened key crop plants, such as the brassicas, for analysis based on dense marker sets and a genomic-sequence backbone. Pests such as the cotton bollworm can be examined genomically. Even challenging poly- and aneuploid genomes such as wheat or sugar cane are now feasible. With food supply under climate change and population pressures an increasing concern, with the agricultural sector of Australia still a strong driver for economic growth, and with Australia's track record in translating scientific developments into practice, there is great opportunity not only to participate in international pre-competitive genomics research, but also to apply the results to improve the national competitive position and create greater global security in food supply. The bottleneck in the application of this information is the lack of bioinformatic infrastructure, particularly personnel.

Case Study 3: Biodiversity and evolution

Australia has been blessed with unique fauna and flora that can provide insights into evolution across all of life. A high-profile example is the extensive resource of marsupials and monotremes in Australia. Using the new technology, researchers have generated the complete sequence of the Tammar wallaby genome, and are considering a project for the echidna genome. This research is of global interest because these Australian species tell us about mammalian evolution in general. Again, the bottleneck in these projects is in bioinformatics, in particular in access to skilled personnel.

These are just three among many Australian research efforts that suffer from a critical lack of bioinformatics support. There is a tremendous opportunity that removing this bottleneck across the biological sciences in Australia would lead very quickly to a far greater return on existing investments, and to a stronger position for Australia in international science and knowledge-based industries.

The opportunity for catalytic change

A variety of approaches could be considered for investment in bioinformatics, coupled with appropriate national coordination to ensure its best use. The over-arching goal would be to ensure that the Australian biosciences R&D community has merit-based access to a skilled bioinformatics team. These bioinformaticians must sit in a local and national context to make their work as productive as possible.

Sketched below is a feasible plan with broad support that would provide this investment and coordination.

Joint local + national investment in bioinformatics infrastructure

As there is already a good national system of computational infrastructure, the major missing component in bioinformatics infrastructure is appropriately skilled personnel. I suggest a co-investment from a national scheme with local (typically State-based) components to develop or consolidate critical mass in each local area while strengthening linkage between locales. These local components will be most effective when located in close association with significant biological science investment. One of the local areas should have double the capacity, allowing it to be the national bioinformatics "lead" and coordinate activity. During visits to major locations around Australia, I found considerable local initiatives to coordinate local (institutional and State

government) funds and to improve local coordination; a national co-investment framework would support these existing efforts, and provide impetus elsewhere.

The goal should be to have around 20 dedicated bioinformatics personnel at each site in steady state (40 in the national lead site), with 50% of those supported by national funds and 50% from local schemes over the first five years. This investment should be nationally coordinated to ensure productivity and collaboration. Each scheme would have to show how this resource could be utilised effectively by all local scientists and how it would directly impact their work. It would be inappropriate to require the same organisational structure in every location; however, a critical peer-review processes should assess the suitability of any proposed structure, placing clear emphasis on management competence, scientific leadership in bioinformatics, and appropriate governance with both local and national components. A particular concern would be the assignment of in-kind local costs from existing funding streams, which might be presented as local investment but in fact are not coordinated in a way that provides critical mass in the local area. Any unclear governance of the local investment, or unclear geographic spread (thus preventing a critical mass of local bioinformatics skills), should be heavily penalised in the peer-review process.

One would budget for a progressive ramp to 20 bioinformaticians per site, so the first two years would show a ramp in hiring. In subsequent rounds beyond the initial five years, one would have to consider strategically the need for this “capacity funding” nationally, and the individual success or otherwise of each local scheme. Local schemes would be expected to show clear synergies with existing investments in biological and computational infrastructure, and should have a clear governance structure which handles potential conflicts or prioritisation issues locally and can contribute to the national system. Nationally the maturity of the field might be such that such locally coordinated capacity is not required beyond five years, or alternatively another capacity round might be appropriate. I recommend considering this topic four years into a five-year scheme, with the explicit expectation that any second round of national-led funding in a future five-year period should demonstrate a transition process to a long-term steady-state structure. The latter might be more-traditional response-mode research funding.

National coordination of bioinformatics infrastructure

A critical component in the successful deployment of a bioinformatics infrastructure will be coordination and synergies across Australia. It would not serve Australia to indulge in needless inter-group direct competition; all of life science is moving to a predominantly international axis, and there is so much more to gain by collaboration and synergy in these pre-competitive areas than the small gain in heavily competitive behaviour. To provide national coordination, one site should have twice the funding level of the other sites, providing a clear centre of gravity for bioinformatics in Australia. This site would have the responsibility to provide open national-level access to large-scale international data resources and coordination of (for example) archiving efforts, in each case being mindful of the international context. Training resources would also be coordinated from this site. The location of this site should be chosen by rigorous peer-review.

It is also critical that this bioinformatics infrastructure be responsive to the needs of biological scientists, whilst being a (large) collective client to the well-managed compute and disk infrastructure already present in Australia. I propose that Australia leverage the nascent national molecular biology framework founded by EMBL Australia, with a specific bioinformatics structure formed as part of EMBL Australia. This would require EMBL Australia to have a clear national role, and to provide an oversight process that considers all of Australia's needs. It would have input from eResearch, computational and other infrastructures, and scientific leadership and representation from funding agencies and key research institutions. Placing the bioinformatics infrastructure in the context of practicing, high-level molecular biology expertise—such as that provided by EMBL Australia—is critical to its success in delivering to a science-driven agenda. This National Bioinformatics Consortium would provide:

- (a) a yearly national coordination activity (likely to include a conference) among sites to ensure coordination and information flow,

- (b) the ability to provide specific funding to stimulate information flow, e.g. short- to medium-term exchange of personnel,
- (c) coordinated training schemes for end-users of bioinformatics, and
- (d) an appropriate peer-review of the local schemes' success in operating the infrastructure, to ensure that any underperforming component can recognise and rectify any issues during the five years.

Indicative costings

Indicative costings for this infrastructure outlined below may be categorised into:

technical personnel:

assuming 4 local areas at 20 individuals each, 10 of which nationally funded; 1 area at 40 individuals, 20 of which nationally funded; and a starting ramp process of 40% in year 1, 70% in year 2 and 100% in years 3–5, split 50:50 national vs local throughout

additional costs for small-scale compute infrastructure.

Exchange program:

Support for twenty 2–6 month exchanges per year (10 episodes in the first year) @ \$20K each including accommodation and travel = \$ 1.8M over five years

National coordination function

2 administration persons * 5 years = \$ 1M over five years

Annual National Conference

Support for a 200-person national conference each year, including international invited speakers and student travel subsidy

Outreach to secondary schools and the public:

1 coordinator + travel

National Funding required	year 1	year 2	year 3	year 4	year 5	total
inflation index	1.000	1.030	1.061	1.093	1.126	
technical personnel (bioinformatics support) *	1,200,000	2,163,000	3,182,700	3,278,181	3,376,526	\$13,200,407
<i>local area staffing (FTE)</i>	4	7	10	10	10	
<i>salary costs (assume \$80k pa in year 1 per FTE)</i>	320,000	576,800	848,720	874,182	900,407	
<i>employment costs (assume 25%)</i>	80,000	144,200	212,180	218,545	225,102	
<i>bioinformatics hub staffing (FTE)</i>	8	14	20	20	20	
<i>salary costs</i>	640,000	1,153,600	1,697,440	1,748,363	1,800,814	
<i>employment costs (assume 25%)</i>	160,000	288,400	424,360	437,091	450,204	
national coordination	192,500	198,275	204,223	210,350	216,660	\$1,022,009
<i>project officers (2 FTE)</i>	150,000	154,500	159,135	163,909	168,826	
<i>employment costs (assume 25%)</i>	37,500	38,625	39,784	40,977	42,207	
<i>miscellaneous costs</i>	5,000	5,150	5,305	5,464	5,628	
annual conference support	80,000	82,400	84,872	87,418	90,041	\$424,731
<i>conference expenses</i>	80,000	82,400	84,872	87,418	90,041	
outreach program	95,000	97,850	100,786	103,809	106,923	\$504,368
<i>coordinator</i>	75,000	77,250	79,568	81,955	84,413	
<i>travel expenses</i>	15,000	15,450	15,914	16,391	16,883	
<i>misc costs</i>	5,000	5,150	5,305	5,464	5,628	

* equivalent funding provided locally

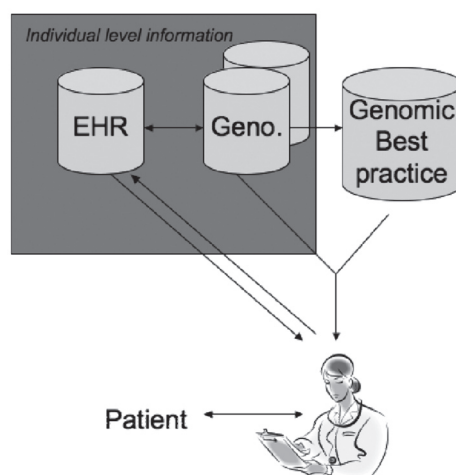
White paper on the informatics outlook for genomic medicine.

The landscape for human genetic research and clinical application.

Over the last 2 years there has been a revolution in human genetics due to the ability to get well powered genome-wide case/control studies executed on human disease. This methodology is very broadly applicable and can uncover quite subtle effects – a leading group in this area is the Wellcome Trust Case Control Consortium, which published in 2007 the analysis of 7 diseases. This has led to direct discoveries and many follow up studies into these diseases.

It is important to stress that although this research is tremendously exciting, it is not yet affecting mainstream clinical practice. Different human genetics researchers have different opinions on the best route to translate the findings of these studies into clinical practice, and more fundamentally there is a vibrant debate on the make up of genetic disease – in brief, the observed genetic risk component discovered by these methods do not explain the far larger family based heritability measures, meaning that there is some additional process occurring. A variety of aspects are candidates for this mismatch – a larger contribution from different rare variants, genetic epistasis interactions between variants on different loci and gene/environment interactions are all plausible aspects of this. There are already some promising signs in leveraging genetics in a clinical setting, for example, severe adverse reactions to high statin levels or using genetic tests to provide an earlier prediction of the correct warfarin dose.

Although there is considerable uncertainty in these features, it seems very likely that the use of genetic information to improve clinical practice will occur for at least 1 common disease within the next 10 years, and seems quite likely that it will be appropriate for more than 1 disease. Genome-wide genotyping costs – currently around \$300 and likely to continue to drop – will mean the cost of the genetic test will be reasonable. This should be seen in the general context of more molecular based diagnostic tools being deployed across the health care arena, though the fact that genomic information can be measured once and reused throughout a patient's life means there are different dynamics to the use of this information. In addition, in this time frame it may be cost effective to provide a more sequencing based approach (costs for full genome resequencing is around \$200,000 currently, down from the mid-millions, but the technology development is continuing to drop this price aggressively). Importantly for this discussion one should be planning on the basis of genetic information being needed by practicing, point of care clinicians within 10 years.



The features of genome-wide genetic data.

There are a number of features of genome-wide genetic data which makes its use quite different from previous medically relevant datasets. Figure 1 shows a possible outline of how one might manage this information and the basic flow of information through the doctor to a patient. The grey box indicates individual data, which includes both updated electronic health care information and genotype information which will be rarely updated once determined. These two datasets would be linked, as would other large scale complex data (shown as another database behind genotypes) such as, for example, MRI scans. Although at one level

genotyping/sequence data is “just another” molecular diagnostic dataset, aspects of this information are different from other medical datasets/

1. The information is applicable to an individual throughout his or her life, and not just for a particular health-care episode. Of course, each particular health care episode is likely to be handled by a unique constellation of professionals who will draw on a variety of information sources, of which the genome will be only one.
2. The information is likely to be informative to a variety of clinicians, often not in the same sub-discipline
3. The information can be used for future, currently not understood clinical uses
4. The information rests on a global information infrastructure about our knowledge and understanding of the human genome which will be continually refined and updated over the next 50 years.
5. The information (genotype or sequencing) is non-trivial to store and manage

Other clinically relevant datasets have some of these features – for example, MRI scans are non-trivial to store and manage, and a patient’s health record is relevant to a variety of clinicians, but the combination of the broad clinical use of this information coupled with its longevity across multiple health related episodes and its complex interplay with genomics research makes this quite a challenge to craft the right information infrastructure for this.

Components of a successful clinical genetic infrastructure

There are technical, management, clinical practice and legislative components of a genetic infrastructure. The following components are required:

Technical

1. A robust, well implemented system which can store the information for many individuals over their lifetimes and provide that information to relevant health professionals, and this information should be a component of an individual’s broader electronic health care record
2. A secure system such that only appropriate people can view information
3. A system with an understanding of the underlying human genome knowledgebase and the ability to update appropriately as new knowledge is incorporated

IT Management

1. That this structure and information can be used across all sub-disciplines in medicine
2. The creation of productive links between the clinical practice database, data flow and information and both the clinical research and human genetics/genomics community.
3. An increase in the capacity and leadership skills in informatics (the individuals who run these technical systems) in a clinical setting. –

Clinical practice and Legislation

I am not the best person to comment on this. In meetings about this in which I have attended in both the UK, comments have included the curriculum for medicine including more genetics at a foundational level and for retraining currently qualified health care professionals. In addition there will clearly be many specific interactions on the use of genetic information in a particular setting – for example, for patient presenting with problem X, what are the critical pieces of information to be aware of. In the legislative area, comments have included the oversight for access of genetic information, handling the risk of accidental release of information and handling the use of genetic information by health care, insurance companies and individuals.

Conclusions.

When one considers the tasks to set up this infrastructure and its interactions with different components of the health care system and society at large it is quite daunting. Clearly this will be a large IT project in a complex area (health care) with complex requirements with research. It is tempting to either postpone discussions due to the current uncertainty of the details of genetic research (alluded to above) or to focus on one or two key components (often the more societal ones are interesting to explore). However, on a practical planning basis, it is critical that this is thought about now, as when the use of genetic information can really impact at least one disease it will be critical that this is implemented in a sensible way. To prepare for that one needs to start planning the system now with all these inter-locking components.

At the EBI we see one of our key missions over the coming decade to be a strong component in the provision of stable, useful information on the human genome, and provide training and expertise transfer to medical informatics components in our member states. It is extremely clear that it is not appropriate for the EBI to be directly implementing these components for practicing healthcare – this must be done in the context of each health care system and in the context of each country – but we are already participating in the large clinical research studies which underpin this information (for example, the Wellcome Trust Case Control Consortium data is housed at the European Genotype Archive – EGA). In terms of information architecture this will have many of the same issues as the member-state implementation process. Our goal is to enable both research in this area, the translation of research to clinical practice and the transfer of expertise in information management to the member states.

White paper on the use of new sequencing technologies and their impact on bioinformatics.

New sequencing technologies

The last four years has seen a revolution in sequencing technology with 3 new high throughput technologies coming on the market – pyrosequencing (Roche/454), sequencing by synthesis on a glass-slide structure (Illumina/Solexa) and sequencing by ligation on a glass slide structure (ABI Solid). The last two have dramatically shorter read lengths to the older capillary technologies (36-50bp) but at far, far higher read numbers (in the billions). The 454 technology is more comparable to the old style technology, but at about 20-fold lower cost.

The cost point of generating sequence is dramatically lower – 1,000 fold in some scenarios and this is continuing to drop. This incredible improvement of cost (indeed, sequencing is one of the few technologies which year-on-year outperforms the increase of computers described as Moore's law) has opened up huge new areas of applications for these machines. It has also significantly shifted the science towards informatics, with the need for both supporting informatics directly on the machines and analysis informatics for the data processing.

This white paper explores a little the new application areas for these machines, and then outlines the informatics challenges.

Basic areas of application

There are a variety of ways of classifying the application areas for these new machines – I will focus mainly on a breakdown by experiment-type, with then the biological area where this can be applied.

De Novo sequencing. Around 2 years ago it was considered unfeasible to use such short reads for de novo genome assembly; however a variety of groups (including Dr Birney's at the EBI) have shown in both simulation and real data that one can use these short reads when done in read-pair format to produce large, draft-quality assemblies. The consumables cost ranges from under \$1,000 for a bacterial genome to around \$10,000 for a 50MB genome, at around 8MB non-redundant sequence. In theory there is no practical limit on the assembly target size, though the engineering challenges for large assemblies are considerable (but not unfeasible). For example, the wheat genome would be accessible using this new methodology. This can be used in all areas of biology; for the sampling of other close primates for human evolution, for human, animal and plant pathogens, for species of importance for environmental science and species for importance of evolutionary positions.

Resequencing for variants. This is the application which stimulated the development of these machines and has a clear impact on human genetics. Large scale resequencing is feasible for any species with an assembly at increasing lower cost. In fact the difference in cost between resequencing and de novo assembly is closing (and indeed the informatics approaches are also converging) but by using population genetics approaches, very cost effective methods can be constructed for getting a very accurate representation of all shared variants. This is being led by the 1,000 genomes project. Again, the EBI has extensive experience in this area, being the co-lead on the data flow for the 1,000 genomes project (Dr Paul Flicek)

Sequencing for an assay readout. The low cost point of these new machines allows radically new assay methods based on sequencing rather than hybridisation. This spans all of molecular biology that can create some sort of DNA readout; transcriptomics, Chromatin immunoprecipitation methods (Chip), methylation and many more besides. This has been a rather unexpected growth area, but makes this machine applicable to nearly every molecular biologist in some way. One leading project in this area is ENCODE, in which Dr Birney and Dr Flicek from the EBI both participate (Dr Birney coordinates the analysis for ENCODE).

Informatics requirements

The low cost of data from these machines has switched much of the science and cost into informatics – in effect, if one can work out how to get some assay or experiment on these machines, then one can immediately scale over the entire genome and often over many samples – but the bottleneck is in data processing and analysis. The current rule of thumb is that the informatics cost is at least as much as the consumables cost, and often far higher. It also has challenging disk requirements. There are three components to the informatics.

Infrastructure Bioinformatics. This is an important task which must be coordinated internationally, for example, the coordination of the human genome sequence, or collating and organising all the protein structures. Exemplar institutes in this area are NCBI and EBI, and in most countries there will be infrastructure activities that feed into these large centres, but unlikely to be a large independent institute dedicated to infrastructure bioinformatics. However, many of the informaticians in the other areas should have good relationships and/or work as part of Infrastructure bioinformatics.

Production informatics. This is the informatics directly associated with the machine. For software processing of the data, each machine needs considerable disk space (this is sometimes bundled with the machine) and the production informatics needs a well provisions set up with disk to store datasets in a variety of transformed states before handing them on. The production compute centre must be close to the physical machines with a good network infrastructure. Usually the cost for production informatics is best handled as a straight component of the machine (ie, in the price for getting a run).

Analysis informatics. This is the informatics associated with the analysis of the data. For some applications this is becoming more standard, though often still very challenging to execute but for most cases there is considerable variability and innovation in the analysis, matching usually the variability and innovation in the sample preparation. Nearly all applications are demanding in terms of disk (again, multiple terabytes of scratch space required) and some demanding in terms of compute and physical memory machine size. In theory this component should occur close to sample generation and the biology drivers of the project. However, in most institutes and in general in nearly all countries there is a severe shortage of analysis informatics capacity, making this often the main bottleneck in analysis.

Addressing this capacity shortfall is complex – over the long term one needs to have PIs comfortable with both “wet” and “dry” components of their science, and hiring at least as many dry individuals as wet. This implies in the undergraduate and graduate training continuing the shift towards “dry” biology, and also increasing the amount of quantitative skills around. This is a hard process to execute in any country. In the short term there is often an under appreciation by PIs from a wet background of the analysis challenges here, with sometimes completely inappropriately provisioned systems. Even with the goals for a well provisioned and staffed centres, attracting the right individual with appropriate experience is hard as they are in short supply. Exploring the different options for how to best handle this short term “analysis crunch” would be very relevant.

The EBI, as one of the largest informatics institutes worldwide, is a leader in all these areas, and there should be productive ways to leverage Australia’s relationship with the EBI. This includes transfer of expertise and deepening training routes, probably via personnel visiting on 3, 6 or 12 months stays at the EBI. Specifically in the area of genomes, the EBI’s genome resources (Ensembl and Ensembl Genomes) proactively work with communities around sets of genomes to improve their information infrastructure. We do this by usually getting joint funding with the community in a particular region. In the context of genomes of particular relevance to Australia (ranging from the Wallaby to Wheat) we would be happy to explore different options.



Brief Biographies

Professor Ewan Birney

Dr Ewan Birney is a Senior Scientist at EMBL (European Molecular Biology Laboratory) working at the EBI (European Bioinformatics Institute). He is better known as one of the founders of Ensembl, and the main Principle Investigator at the EBI until 2007.

Professor Nadia Rosenthal

Acclaimed researcher Professor Nadia Rosenthal leads the Australian Regenerative Medicine Institute (ARMI) at Monash. Her research concentrates on embryonic heart development, ageing mechanisms and stem cell-driven regeneration of neuromuscular and cardiac tissue, using the mouse as a model for human response to disease.

Professor Rosenthal has exceptional scientific credentials, including sixteen years working at Harvard Medical School. She currently directs the influential European Molecular Biology Laboratory (EMBL) Outstation in Monterotondo, Italy, one of five EMBL campuses with over 1500 employees across Europe.

She also serves as scientific director of the Heart Science Centre at Imperial College London and is currently working with leading heart transplant surgeon Professor Sir Magdi Yacoub on developing new regenerative ways to treat heart failure.

Comprehensive List of Participants

Abbott, Cathy	Flinders University
Aitken, Alex	Department Innovation, Industry, Science and Research
Applebe, Bill, Dr	Victorian Partnership for Advanced Computing
Arthur, Jonathan, Dr	CEO/Director Sydney Bioinformatics, University of Sydney
Bacic, Tony, Prof	University of Melbourne
Barendse, Bill, Dr	Molecular Genetics CSIRO Livestock Industries
Barker, Jeremy	CEO, Queensland Facility for Advanced Bioinformatics
Bellgard, Matthew, Prof	Centre for Comparative Genomics, Murdoch University
Belov, Kathy, Dr	Australasian Wildlife Genomics Group, University of Sydney
Black, Steve	Victorian e-Research Strategic Initiative
Bonnington, Paul, Prof	Monash e-Research Centre
Borda, Ann, Dr	Victorian e-Research Strategic Initiative
Bowen, Joanne	NHMRC Research Fellow, Mucositis Research Laboratory, IMVS
Bowtell, David, Prof	Peter Mac Research
Brown, Margaret	Consultant, Minter Ellison
Buckle, Ashley, A/Prof	Monash University
Burden, Conrad	ANU Centre for Bioinformation Science
Bushell, Gillian, Prof	Dean (Research)
	Science, Environment, Engineering and Technology, Griffith University
Carroll, Jacqueline	Directorate Manager, Molecular Pathology, IMVS
Carter, Kim, Dr	The Telethon Institute for Child Health Research
Chalk, Alistair, Dr	Research Fellow, Systems Biology Program, Eskitis Institute, Griffith
Chen, Phoebe, A/Prof	Deakin University
Chetty, Madhu, Dr	Monash University
Clark, Georgina, Dr	Immunoregulation Team Leader, Mater Medical Research Institute
	Medical Research Institutes
Cochrane, Tom, Prof	Deputy Vice-Chancellor, Queensland University of Technology
Cocks, Ben, Dr	Department of Primary Industries
Coddington, Paul	Deputy Director, eResearch SA
Collins, Grant	ADFA Chemistry
Conway, Topaz, Dr	Deputy Director (Development)
	Mater Medical Research Institute, Medical Research Institutes
Coppel, Ross, Prof	Monash University
Crowe, Mark, Dr	Acting Production Manager, Pfizer Animal Genetics
Curtin, Chris	Senior Research Scientist, Australian Wine Research Institute
Cuthbert, Brad	Principal Project Officer
	Office of Biotechnology, DTRDI, Queensland Government
Dalrymple, Brian, Dr	Science Leader Bioinformatics, CSIRO Livestock Industries
Datta, Amitava, Prof	University of Western Australia, School of Computer Science and Software
	Engineering (CSSE)
Dawes, Ian, Prof	Director, Ramaciotti Centre for Gene Function Analysis,
	University of New South Wales
De Kerk, Nick, Prof	Telethon Institute for Child Health Research
Dinger, Marcel, Dr	Genomics and Computational Biology, The University of Queensland
East, Peter	CSIRO Entomology
Easteal, Simon	ANU JCSMR
Edwards, Dave, Dr	Bioinformatics Leader, Australian Centre for Plant Functional Genomics,
	University of Queensland
Edwards, Owain	CSIRO Entomology
Fisk, Nick, Prof	Director, Centre for Clinical Research, The University of Queensland
Forrest, Sue, Dr	Australian Genome Research Facility
Francis, Rhys, Dr	Platforms for Collaboration
Gecz, Josef	Professor (Human Genetics), University of Adelaide
Gilbert, Andrew	Bioplatforms Australia
Giles, Keith, Dr	University of Western Australia, Centre for Medical Research, WAIMR



Gonda, Tom, Prof	Director of Research Diamentina Institute for Cancer, The University of Queensland
Goodall, Greg	Associate Professor Department of Medicine/Head, Cytokine Research Laboratory, IMVS CSIRO Entomology
Gordon, Karl	Technical Manager, Queensland Facility for Advanced Bioinformatics
Gorse, Dominique	ANU RSBS
Graves, Jenny	Manager, Technology and Commercialisation, DTRDI, Qld Government
Gray, Jill, Dr	Manager, Emerging Technologies Queensland Dept of Primary Industries and Fisheries
Grieve, Paul, Dr	Genomics and Computational Biology, The University of Queensland
Grimmond, Sean, A/Prof	Technical Officer Microbial Metabolomics, Australian Wine Research Institute
Hack, Jeremy	Principal Research Scientist and e-Health Theme Leader, Australian e-Health Research Centre, CSIRO
Hansen, David, Dr	School of Software Engineering and Data Communications, and Microsoft e-Research Centre, Queensland University of Technology
Hogan, James, A/Prof	ANU JCSMR Monash University
Huttley, Gavin	Director, Office of Biotechnology, DTRDI, Queensland Government
Irving, James, Dr	University of Sydney
Jacobs, Mark, Dr	Executive Dean, Faculty of Science and Technology, QUT
Jermiin, Lars	Diversity Arrays Technology Pty Ltd
Kaplan, Simon, Prof	University of Melbourne
Kilian, Andrzej	La Trobe University
Konagurthu, Arun, Dr	Manager Australian Institute for Bioengineering and Nanotechnology, UQld
Kosten, Mark, Dr	University of Western Australia, Centre for Medical Research, WAIMR
Kromer, Jens, Dr	Chair, The Western Australia Institute for Medical Research University of Western Australia, Centre for Medical Research
Laing, Nigel, Prof	Monash University
Leedman, Peter	Associate Professor, School of Computer and Information Science, UniSA
Lefevre, Christophe, Dr	CSIRO Mathematical and Information Sciences
Li, Jiuyong	University of Western Australia, DVCRI
Lovell, David	Australian Genome Research Facility
McEachern, Doug, Prof	University of Canberra – Forensics
McGrath, Annette, Dr	Coordinator, Microbial Metabolomics, The Australian Wine Research Institute
McNevin, Dennis	ARC Centre of Excellence, Plant Energy Biology
Mercurio, Meagan	Head of Biological Sciences, Flinders University
Millar, Harvey, Prof	Senior Director, Office of Health and Medical Research
Mitchell, Jim	Queensland Health, Queensland Government
Mortimer, Robin	Principal Policy Officer, Science and Innovation Directorate, DFEEST
Murrell, Ken	Pangalax
O'Callaghan, John, Prof	CSIRO Entomology
Oakeshott, John	CEO, Queensland Cyber Infrastructure Foundation
Pailthorpe, Bernard, Prof	ANU Biomolecular Resource Facility
Palmer, Stephanie	Walter and Eliza Hall Institute of Medical Research
Papenfuss, Tony, Dr	University of Melbourne
Rathjen, Peter, Prof	Head, Genomics and Computational Biology, The University of Qld
Ragan, Mark, Prof	Director, ARC Centre of Excellence in Bioinformatics University of Western Australia, Centre for Medical Research, WAIMR
Ravine, David, Prof	e-Science Program, Faculty of Information Technology, QUT
Raymond, Kerry, Prof	Bioinformatics, CSIRO Livestock Industries
Reverter-Gomez, Antonio, Dr	Executive Director Technology and Emerging Industries, Queensland Government
Roach, Leigh	

Roberts, Claire	Associate Professor Research Centre for Reproductive Health, University of Adelaide
Rodgers, Ray	Research Centre for Reproductive Health, University of Adelaide
Rosenthal, Nadia, Prof	Director, Australian Regenerative Medicine Institute Director, European Molecular Biological Laboratories
Rudzki, Zbigniew	Head, Department of Molecular Pathology, IMVS
Russell, Paul	Executive Director Enabling Technologies, DTRDI, Queensland Government
Ryan, Louise	CSIRO Mathematical and Information Sciences
Saint, Rob	ANU RSBS
Sargent, Mike	MA Sargent and Associates
Saunders, Ian	CSIRO Mathematical and Information Sciences
Savin, Keith, Dr	Department of Primary Industries
Sawbridge, Tim, Dr	Department of Primary Industries
Scott, Paul, Dr	ARC Centre of Excellence for Integrative Legume Research The University of Queensland
Shannon, Elizabeth	Manager, ICT, Industry Development, DTRDI, Queensland Government
Siddle, David, Prof	Deputy Vice-Chancellor, The University of Queensland
Sloggett, Clare, Dr	Intersect Pty Ltd
Smith, Ian, Prof	Monash University
Smyth, Gordon, Dr	Walter and Eliza Hall Institute of Medical Research
Solomon, Patty	Chair, Bioinformatics, University of Adelaide
Speed, Terry, Prof	Walter and Eliza Hall Institute of Medical Research
Stanley, Andrew	Policy and Inter-Government Relations Division Department of Health (DoH)
Stefanov, Valery, Prof	University of Western Australia, School of Mathematics and Statistics
Sterling, Leon, Prof	University of Melbourne
Stowe, Jenny, Prof	Director of Research Institute for Molecular Bioscience, The University of Queensland
Szubert, Marek	IMVS Bioinformatician
Taylor, Geoff, Prof	University of Melbourne
Taylor, Jen	CSIRO Plant Industry
Thomas, Mervyn, Dr	CEO, Emphron Informatics
Thomas, Torsten, Dr	School of Biotechnology and Biomolecular Science, University of New South Wales
Tizard, James	Director (Interim), eResearch SA/CEO, SABREnet
Tizjani, Silvio	CEO, Australian Regenerative Medicine Institute
Tonti-Filippini, Julian	ARC Centre of Excellence, Plant Energy Biology
van Niekerk, Alvin, Dr	COO, ARC Centre of Excellence for Integrative Legume Research, The University of Queensland
Viney, Geoff	Department of Primary Industries
Wade, Clare, Prof	Professor of Animal Genetics and Computational Biology, University of Sydney
Wakefield, Matthew, Dr	Walter and Eliza Hall Institute of Medical Research
Wallace, Iain, Prof	Victorian Partnership for Advanced Computing
Webb, Geoff, Prof	Monash University
Whisstock, James, Prof	Monash University
Wilkins, Marc, Prof	Director, NSW Systems Biology Initiative, University of New South Wales
Wilkinson, Ross, Dr	Australian National Data Service
Williams, Rohan	ANU JCSMR
Wilson, Sue	ANU Centre for Bioinformation Science
Wong, Lanna	Manager, ARC Centre of Excellence in Bioinformatics, UQld
Yu, Bing, Dr	Acting Director, SUPAMAC, University of Sydney
Zobel, Justin, Prof	University of Melbourne, National ICT Australia

Participants



Secretariat



Printed on Revive Silk using vegetable inks.



australian made



recycled



sustainable forest



PEFC



totally chlorine free



elemental chlorine free