

2006 ANNUAL REPORT

ARC CENTRE IN BIOINFORMATICS



Australian Government
Australian Research Council



CHAIR'S FOREWORD	2		
DIRECTOR'S REVIEW AND HIGHLIGHTS	3		
ORGANISATION AND GOVERNANCE			
Organisational chart	4		
Advisory Board	5		
Research Committee	8		
Director	9		
Chief Investigators	9		
Partner Investigators	11		
Special advisor	12		
Strategic partnerships	12		
Information infrastructure	13		
Centre Manager	13		
Postdoctoral Fellows/Research Officers	13		
Research Assistants	13		
Masters/PhD students	13		
Honours students	14		
International & national interns	15		
Exchange student	15		
Research trainees and volunteers	15		
Visiting scholars, associates	15		
STRATEGIC DIRECTION	17		
Vision	17		
Mission	17		
Goals	17		
Objectives	17		
Capabilities – core technologies and infrastructure	17		
RESEARCH PROGRAMS	18		
Program 1 – Experimental phenomics	19		
1.1 Computational and microarray-based analysis of transcriptional complexity of mammalian cells	19		
1.2 Computational cellular biology: combining computational prediction and experimental validation	20		
1.3 Generation of the experimental data for the modelling of cellular compartments and the development of organelle biogenesis phenotype screens	21		
Program 2 – Modelling cellular networks	22		
2.1 Hybrid algorithms for large-scale network and graph optimisation problems in bioinformatics	22		
2.2 Computational discovery of gene-regulatory motifs and motif clusters	24		
2.3 New statistical approaches in bioinformatics	25		
2.4 Mathematical modelling of gene regulatory networks	26		
2.5 Complex systems modelling of regulatory and protein-protein interaction networks	28		
2.6 Ortholog mapping of prokaryotic and eukaryotic genomes	29		
Program 3 – The Visible Cell™ – Data integration and visualisation	31		
3.1 Advanced data management and integration	31		
3.1.1 Data environment for the Visible Cell™ project	31		
3.1.2 Integrating database technologies and visual analysis for genome data	33		
3.2 Visualisation	35		
3.2.1 Concept and engineering of the Visible Cell™	35		
3.2.2 Spatial modelling of cellular systems	37		
3.3 Bioinformatics web services	39		
KEY PERFORMANCE INDICATORS	41		
1. Research findings	41		
Quality of publications	41		
Number of publications	42		
Patents and trademarks	42		
Invitation to address and participate in conferences	42		
International conferences	42		
National conferences	44		
Invitations to visit leading international research groups and research centres	45		
Commentaries about the Centre's achievements	46		
Evidence of interdisciplinary linkages within the Centre	46		
2. Research training and professional education	48		
Recruitment	48		
2006 Winter School in Mathematical and Computational Biology	48		
ACB-BINZ Bioinformatics Student Symposium	49		
International exchange program	49		
Other research training and professional education	49		
Undergraduate courses in the Centre's area	50		
Senior high school student research scheme	50		
3. International, national and regional links and networks	50		
International visitors	50		
National visitors to one or more ACB nodes	51		
Sabbatical visitors	51		
National and international workshops	51		
Other research linkages	52		
4. End-user links	53		
Commercialisation activities	53		
Government, industry and business briefings	53		
Training in technology transfer and commercialisation	54		
Public awareness programs	54		
5. Organisational support	54		
New organisations recruited to or involved in the Centre	54		
Infrastructure	54		
Cash and In-kind contributions: collaborating organisations	56		
6. National benefit	57		
APPENDIX 1: PUBLICATIONS AND PAPERS	58		
APPENDIX 2: FINANCIAL STATEMENT	66		





The 2006 annual report offers a broad-based view of the ARC's Centre in Bioinformatics (ACB), operating at its full complement of personnel and research projects.

I am very pleased to report to you that, over the past year, the ACB has continued to demonstrate very solid progress in terms of its research outcomes. For example, during the last calendar year, the Centre more than doubled its C1 journal output, while continuing to achieve publication success in top journals.

The ACB team is a highly collaborative one and involves an integrated research portfolio in bioinformatics with various activities operating at the Universities of Queensland, ANU, Deakin and Newcastle, together with numerous international collaborations.

It is noteworthy that the Centre's Visible Cell™ project gained significant momentum during 2006. This project is a long-term effort with the aim of visualising protein pathways operating from the cell wall to its nucleus. Early results have been very successful in raising broad interest in this project's potential long-term outcomes.

During 2006, the team was able to wrap up several projects, including Project 1.3, Project 3.3, and the former Project 2.5, which it merged into Project 3.2.2. These project closures have added valuable focus to the continuing research program across the Centre.

ACB completed agreements for two commercial projects with ABI and Pfizer in the year. Work on these projects will commence early in 2007.

In addition to its national bioinformatics research role, ACB continued to provide a program of advanced education and training. During 2006, it hosted a Winter School in Mathematical and Computational Biology. In addition, ACB further developed its strong educational and training support role in Australia (and elsewhere) in bioinformatics.

ACB has built on the strengths and commitment of its researchers across the numerous disciplines required to undertake effective bioinformatics research. These skills sets now include strong bases in mathematics, statistics, computer science, information technology and the molecular biosciences.

In the second half of 2006, ACB completed a detailed review of the Centre's research and future plans with the Australian Research Council. Although continuing funding of the Centre by ARC beyond ACB's current term remains unknown at the time of writing, various other forms of financial support for elements of the Centre's research program have emerged during the past year.

It is my view that the ARC Centre for Bioinformatics continues as a strong and developing centre for bioinformatics in Australia. It has many interdisciplinary linkages, across biotechnology, both in Australia and internationally. With its expanding relationships across the Australian biotechnology sector, I believe that ACB will continue to provide a critical role for the Australian research infrastructure. I do hope that you will concur with this view as you review the ACB's annual report.

Dr Gregory R. Smith
Chair
March 2007





2006 was a productive year for the ARC Centre in Bioinformatics. Working with a full complement of personnel, we achieved important outcomes across our core research areas. New collaborations initiated in 2004-2005 across our geographical nodes began to bear fruit in publications, student completions and additional co-investment. We focused and redeployed resources to build further momentum in our Visible Cell™ project. Among our advanced training initiatives, the annual Winter School in Mathematical and Computational Biology attracted participants from 41 institutions. We increased our already proactive role in national and international bioinformatics initiatives.

The ARC Centre in Bioinformatics (ACB) arose from the 2002-2003 Centres of Excellence funding round of the Australian Research Council (ARC). After interview, our bid for an ARC Centre of Excellence for Genome-Phenome Bioinformatics was placed onto a reserve list. In April 2003 we were invited to revise our application for consideration for funding as an ARC Centre, and on 19 August 2003 the then Minister of Education, Science and Training, the Hon. Dr Brendan Nelson MP, announced funding for our (renamed) ARC Centre in Bioinformatics. ACB began operation on 3 November 2003 with the support of our partner institutions, particularly The University of Queensland. The inter-institutional contract required for release of Commonwealth funds into ACB was executed in August 2004, at which point we were able to proceed fully with recruitment and program-building.

A country – or a funding agency – adds value to its research portfolio by supporting a diversity of R&D modalities, some focused on the excellence of single investigators or small teams, others on multi-investigator networks and centres. Our contribution lies firmly in the latter category. ACB will achieve its objectives by delivering integrated programs that synergise the strengths and commitment of

our researchers from multiple disciplines and institutions. Interdisciplinary synergies were critical to securing major research collaborations in 2006. A new collaboration with Pfizer will build on our expertise in data integration and computational modelling, while an agreement with Applied Biosystems (ABI) will focus on high-throughput DNA sequence generation, data management and algorithmics. These two collaborations will get underway in early 2007.

Success must be built on research excellence. In 2006 our investigator teams published 115 peer-reviewed journal articles and conference papers, including 11 in journals with double-digit impact factors. Six students were co-authors on these high-impact papers. Prof. Geoff McLachlan was awarded an ARC Professorial Fellowship, and Dr Sean Grimmond an NHMRC fellowship. Prof. Mike Fellows received a Humboldt Research Award.

ACB played a key role in establishing the Queensland Facility for Advanced Bioinformatics (QFAB), a Smart State initiative of Queensland Government and seven institutional partners. By leveraging the advanced data and computational infrastructure of QFAB, ACB researchers will be able to keep our primary focus on genome-phenome research, while reciprocally QFAB (and through it, nodes of the national collaborative research infrastructure in biomolecular informatics and their clients) will benefit from association with our research and advanced training. At our University of Newcastle node, the Newcastle Bioinformatics Initiative, an ACB strategic partner since 2004, became a founding partner in the new Priority Research Centre for Bioinformatics, Biomarker Discovery and Information-Based Medicine. ACB alumna Dr Lucia Santoso was named Coordinator of the Australian Bioinformatics Network, managed for the Australian government by Bioinformatics Australia.

ACB is committed to advanced education and training in bioinformatics and allied disciplines. Our third annual Winter School in Mathematical and Computational Biology attracted 148 participants, about half of them postgraduate students, representing 29 Australian and 12 overseas universities, research institutions and

companies. Our students and researchers enjoyed a student bioinformatics workshop hosted by Bioinformatics Institute New Zealand under terms of our 2004 memorandum of understanding. We actively supported BioInfoSummer (Canberra), the Bioinformatics Australia meeting held in association with the annual AusBiotech conference (Sydney) and, in collaboration with IMBcom, student training in bio-business and commercialisation (Sunshine Coast).

An important event in the life of ACB was our mid-term review by the ARC. Our investigators, research teams and institutional partners pulled together not only for the written submissions and site visit, but (perhaps more importantly in the longer term) in collectively imagining the future – where we as a research community want to be in five years, and how we collectively might best get there.

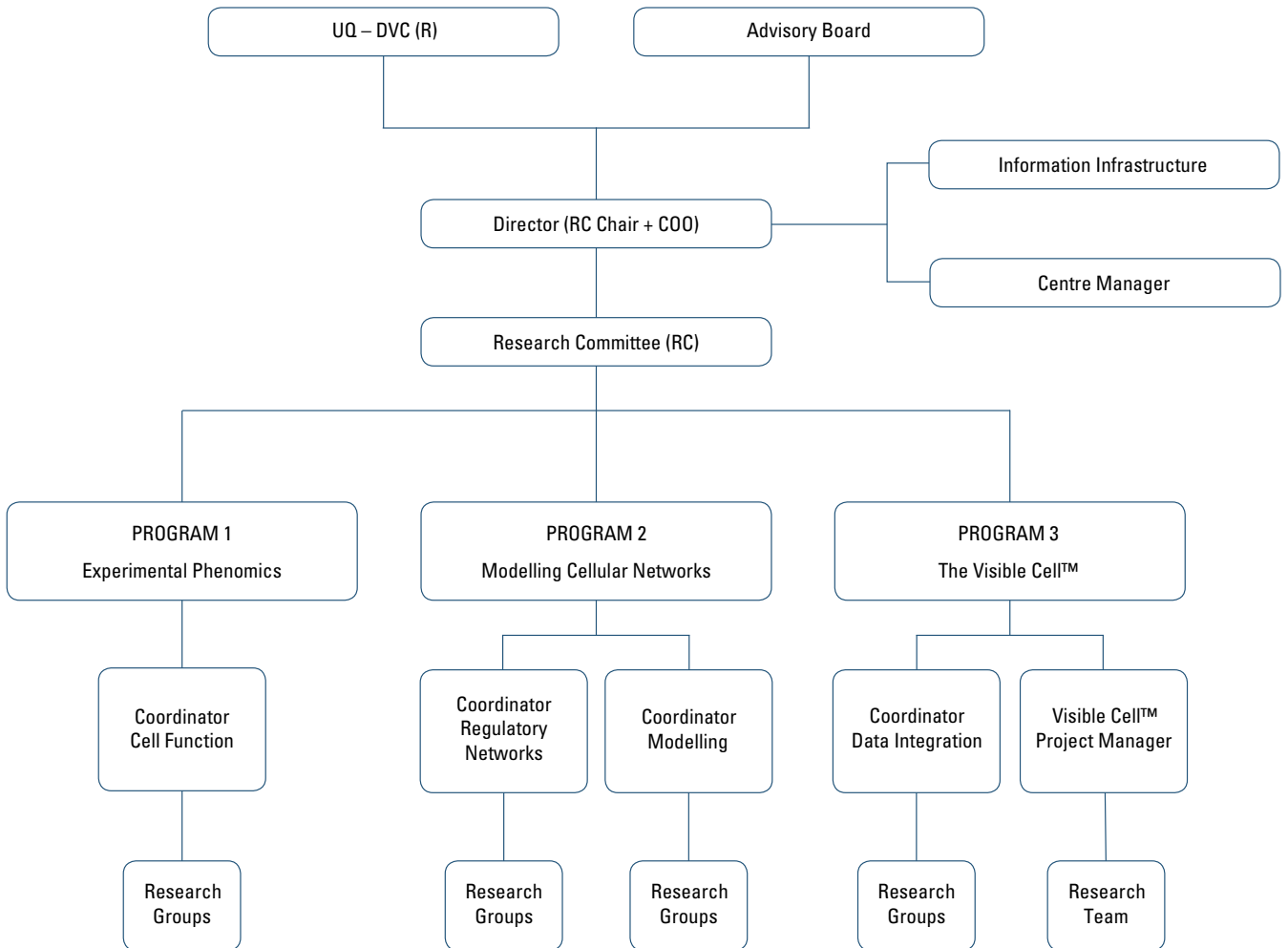
Whatever the outcome, the year ahead – the fifth and final year of our initial funding – will be a pivotal one for bioinformatics in Australia. BioPlatforms Australia (NCRIS 5.1) and its component capabilities have high expectations for our community not just to deliver tools, but to help ensure their effective application to real problems. Through NCRIS the term “e-research” will become highly visible, with bioinformatics its marquee application in the bioscience domain. The National Bioinformatics Network, through Bioinformatics Australia, is mobilising diverse stakeholders in support of the vision inherent in the National Bioinformatics Strategy. Research consortia will increasingly build bioinformatics into the fabric of proposals for new R&D programs and centres. And glancing briefly ahead to 2008, Australia will host the 19th International Conference on Genome Informatics (GIW). The ARC Centre in Bioinformatics looks forward to working with our bioinformatics colleagues and the broader R&D community in these and other important initiatives.

On behalf of my colleagues in the ARC Centre in Bioinformatics, I invite your comment.

Professor Mark Ragan
Director, ARC Centre in Bioinformatics
March 2007



Organisational Chart



Advisory Board

An Advisory Board exercises the highest level of strategic guidance for ACB. Its fourth, fifth and sixth meetings were held on 7 April, 18 July and 6 November 2006. As of 31 December 2006, the members of Advisory Board were:

1 Dr Gregory R. Smith (Chair)

Director
SciVentures Investments Pty Ltd
Melbourne VIC, Australia

Dr Greg Smith is a cofounder and director of SciVentures Investments Pty Ltd, the manager of the SciVentures Pre-Seed Fund. He also is a director of five start-up companies into which SciVentures has invested. Greg has been a member of the Commonwealth Government's Industrial Research & Development Board and of its Advisory Council for Intellectual Property, and continues as Board Chair for the NANO Major National Research Facility. He was awarded the Maurice Holland award of Industrial Research Institute (Washington DC, USA) in 2000 for his "front-end innovation" at AlliedSignal (now Honeywell) and at Alcoa.

2 Prof. Ross Coppel

Director, Victorian Bioinformatics Consortium
Department of Microbiology
Monash University
Melbourne VIC, Australia

Prof. Ross Coppel is a recipient of the Glaxo Award for Advanced Research in Infectious Diseases and of an International Investigator award from the Howard Hughes Medical Institute. He is a named inventor on nine patents for inventions in malaria, primary biliary cirrhosis and novel antibiotics, and is internationally recognised for his work in these fields and in malaria genomics. Ross is Professor of Microbiology at Monash University, Director of the Victorian Bioinformatics Consortium, and co-founder of the biotechnology company Glykoz.

3 Prof. Simon Easteal

Co-Director, Centre for Bioinformation Science
John Curtin School of Medical Research
The Australian National University
Canberra ACT, Australia

Prof. Simon Easteal is a Research Professor at the John Curtin School of Medical Research in the Institute of Advanced Studies, ANU. As head of the Predictive Medicine Group he investigates interactions among genetic and environmental factors in determining illness, personality and human performance, and the predictive value of genetic information in managing health and wellbeing. Simon was founding co-Director of ANU's Centre for Bioinformation Science, Director of the Genetic Epidemiology Unit at the Menzies Centre for Population Health Research in Hobart, and Editor-in-Chief of *Molecular Biology and Evolution*. His work as a Trusted Intermediary in the Australian Industry Group's InnovationXchange Network involved business intermediation, innovation management and advising on R&D commercialisation. He serves on numerous committees and boards in both private and public sectors.

1



2



3





4

4 Prof. John Mattick AO

Professor of Molecular Biology and
ARC Federation Fellow
Institute for Molecular Bioscience
The University of Queensland
Brisbane QLD, Australia

Prof. John Mattick stepped down as Director of the Institute for Molecular Bioscience at the end of 2005 to take up an ARC Federation Fellowship, after serving as Foundation Co-Director (together with Prof. Peter Andrews) from 2000-2002, and as Director from 2003-2005. He was Foundation Director of the Centre for Molecular and Cellular Biology (previously the Centre for Molecular Biology and Biotechnology) from 1988-1999, Foundation Director of the Australian Genome Research Facility from 1996-2002, and Foundation Director of the ARC Special Research Centre for Functional and Applied Genomics from 2000-2002. John's main research interest is the role of non-protein-coding RNA in the evolution and development of humans and other complex organisms. He has received the 1989 Pharmacia-LKB Biotechnology Medal by the Australian Biochemical Society, an Honorary Fellowship of the Royal College of Pathologists of Australasia (2002), the Centenary Medal by the Australian Government (2003), and the CSIRO Eureka Prize for Leadership in Science (2006). He was appointed an Officer in the Order of Australia (AO) in 2001 for service to molecular biology, and biotechnology. Professor Mattick serves on the advisory boards of several institutes in Australia and abroad.



5

5 Prof. Mark Ragan FLS

Director, ARC Centre in Bioinformatics
(*ex officio*)
Institute for Molecular Bioscience
The University of Queensland
Brisbane QLD, Australia

Prof. Mark Ragan joined IMB in 2000 after 28 years with National Research Council Canada, where he co-founded and developed programs in bioactive compounds, molecular biology, genomics and bioinformatics, including Canadian Bioinformatics Resource. His research record includes many "firsts" for his Institute and for NRC, including first use of PCR, first Unix workstation, and first 100 Kbp and 1 Mbp of DNA sequence. With Prof. W. Ford Doolittle and Dr Robert Charlebois he received the largest single grant under the first Canadian genome program (CGAT) and sequenced *Sulfolobus solfataricus*. Mark has been President of two international scientific societies, and a Fellow of the Canadian Institute for Advanced Research. He represents Australia on the executive of the International Conference on Bioinformatics and the Association of Asian Societies of Bioinformatics.

6 Prof. David Siddle

Deputy Vice-Chancellor (Research)
The University of Queensland
Brisbane QLD, Australia

Prof. David Siddle was appointed Deputy Vice-Chancellor (Research) in 2002. He is responsible for enhancement of The University of Queensland's research and research training profile, and development of research collaborations. Previously David was University of Queensland Pro-Vice-Chancellor (Research) 2001; University of Sydney Pro-Vice-Chancellor (Research) 1997-2001; and University of Queensland Dean, Postgraduate Studies 1993-1997. He was chair of the ARC Social Sciences and Humanities Panel 1993-1994, and deputy chair of the ARC Research Grants Committee 1994. He is also a Director of AHURI Queensland Research Centre Ltd; CRC Beef Genetic Technologies; and CRC Mining.



6

7 Prof. Lap-Chee Tsui FRS

Vice-Chancellor
Hong Kong University
Hong Kong

Prof. Lap-Chee Tsui received international acclaim in 1989 when he identified the defective gene (Cystic Fibrosis Transmembrane Regulator, CFTR) that causes cystic fibrosis. His discovery is seen as one of the key breakthroughs in the genetics of human disease. Lap-Chee has received numerous awards and honours, including Distinguished Scientist of the Medical Research Council of Canada, and the Killam Prize. He is a Fellow of the Royal Society of Canada, a Fellow of the Royal Society of London, a Fellow of Academia Sinica, an Honorary Member of World Innovation Foundation, a Foreign Associate of the National Academy of Sciences USA, and an Honorary Fellow of the Royal College of Physicians of London. He has been awarded the Order of Canada (Officer), and the Order of Ontario.



7



8

8 Prof. Marc Wilkins

Department of Biotechnology and
Biomolecular Sciences
University of New South Wales
Sydney NSW, Australia

Prof. Marc Wilkins developed the concept of the proteome and coined the term. He is author of more than 70 peer-reviewed research and review publications and holds patents pending. He is a Senior Editor of *Proteomics*, and served as a Member of the Australian Government's Expert Task Force for Bioinformatics. In 1999, Marc co-founded the biotechnology company Proteome Systems, where he served as Vice President of Bioinformatics (1999-2004) and Head of Proteomics (2004-2005). This company is Australia's premier proteomics company, employs more than 60 staff and is listed on the Australian Stock Exchange. His research interests are proteomics, bioinformatics and their application to biomedical research.





1



2

Research Committee

Members of the ARC Centre in Bioinformatics Research Committee for 2006 were:

Chair	Prof. Mark Ragan
Advisory Board Member	Prof. John Mattick
Program 1	Dr Rohan Teasdale
Program 2	Prof. Kevin Burrage
Program 3	A/Prof. Phoebe Chen
Program 3	Dr Lindsay Hood (Adjunct member to August 2006)
Program 3	Dr Brad Marsh (Adjunct member)
Secretary	Ms Lanna Wong

1 Dr Brad Marsh

Senior Research Fellow and Group Leader
 Institute for Molecular Bioscience
 The University of Queensland
 Brisbane QLD, Australia

Dr Brad Marsh leads the Beta Cell Structural Biology and Cellular Tomography groups at the Institute for Molecular Bioscience at UQ. His research is focused on understanding how the crucial metabolic hormone insulin is processed, trafficked and released from the pancreatic beta cell. This work will identify how and where defects occur at the subcellular level, and how they contribute to the onset of the diseases commonly known as type 1 and type 2 diabetes. Brad is recognised internationally for his cutting-edge application of large-volume EM tomography to visualise mammalian cell structure in 3D at high resolution. His striking images of subcellular architecture in the insulin-secreting pancreatic beta cell have been featured on the covers of *Proceedings of the National Academy of Sciences USA*, *Proceedings of Microscopy and Microanalysis*, *Traffic and Nucleic Acids Research*. His work and methods have been highlighted in *La Recherche*, *Nature Reviews Molecular Cell Biology*, the National Institutes of Health's highly esteemed biennial publication *NCRH Highlights*, the latest editions of *Molecular Cell Biology* (Lodish *et al.*) and *Essential Cell Biology* (Alberts *et al.*) and most recently in *Australian Life Scientist* and *Australian Biochemist*.

2 Dr Lindsay Hood (to August 2006)

High Performance Computing Manager
 Institute for Molecular Bioscience
 The University of Queensland
 Brisbane QLD, Australia

Dr Lindsay Hood has many years' experience with high performance computing in academia, government service and the private sector, including with the legendary Thinking Machines Inc. (Cambridge, MA). Prior to joining IMB, he was a member of Compaq's global life and material sciences team, which was a significant part of the human genome project at the Sanger Centre and at Celera. Lindsay was responsible for high-performance computing resources at IMB and for ACB, and was the technical lead for the Visible Cell™ Project. He is now Head of Grid Programs for the Australian Partnership in Advanced Computing, based at the APAC National Supercomputing Facility at ANU in Canberra.

Director

1 Prof. Mark Ragan FLS

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

Please refer to page 6 for profile.

Chief Investigators

2 Dr Timothy L. Bailey

Institute for Molecular Bioscience
The University of Queensland

Dr Tim Bailey's major areas of expertise are machine learning, statistical pattern recognition (discovery), and the statistics of sequence alignment scores. He is highly experienced in developing parallel algorithms for massively parallel computer architectures, and in developing and supporting web-based algorithms for bioinformatics. He wrote MAST, MEME and MetaMEME, the standard motif-recognition programs in bioinformatics. Tim's MEME website is used by nearly 1000 researchers per month.

3 Prof. Kevin Burrage ARC Federation Fellow

Advanced Computational Modelling
Centre;
Department of Mathematics;
School of Information Technology &
Electrical Engineering; and
Institute for Molecular Bioscience
The University of Queensland

Prof. Kevin Burrage is founding CEO of the Queensland Parallel Supercomputer Foundation (now Queensland Cyber Infrastructure Foundation), the Advanced Computational Modelling Centre (ACMC) at UQ, and the ViSAC visualisation laboratory. He was awarded a prestigious Federation Fellowship by the Australian Research Council in 2003. Kevin has co-authored more than 150 papers in the fields of computational science, computational biology, mathematical modelling and complex systems. This oeuvre consists of fundamental work on the numerical solution of both ordinary differential equations and stochastic differential equations, and algorithms for linear systems. He is author of a monograph on parallel and sequential methods for ordinary differential equations.

4 A/Prof. Yi-Ping Phoebe Chen

School of Engineering Information
Technology
Deakin University

Associate Professor (Reader) Phoebe Chen is Bioinformatics Group Leader, and Head of Multimedia information retrieval at Deakin University in Melbourne. She was Director of Research (Acting Associate Head of Research) in Deakin's School of Information Technology. Phoebe has been working in bioinformatics since 1997. Prior to that, she conducted significant research in knowledge discovery, information retrieval, database query languages, data visualisation, and data integration systems. Phoebe has written about 100 refereed journal articles and international conference papers, and serves on the editorial boards of *IEEE Transactions on Multimedia*, *Current Bioinformatics*, *Journal of Research and Practice in Information Technology*, *International Journal of Intelligent Computing in Medical Sciences*, and *Image Processing*. Phoebe founded Asia-Pacific Bioinformatics Conference and continues to chair its Steering Committee. She serves on the on the steering committee of Bioinformatics Australia.

5 Prof. Michael R. Fellows

Parameterised Complexity Research
Unit
Office of the Deputy Vice-Chancellor
(Research)
The University of Newcastle

Prof. Mike Fellows is internationally recognised for his foundational work on parameterised complexity, for which he and collaborator Prof. Rod Downey (Victoria University, NZ) were nominated for the Gödel Prize in 2005. They co-authored the foundational papers and the first comprehensive monograph for the field, which is now considered one of the main branches of theoretical computer science concerned with algorithms and complexity. For more than a decade, this area has had strong applications in bioinformatics and computational biology, both in the design of effective algorithms for large datasets, and in the formation and evaluation of theory for natural computing systems. Mike is also recognised for major contributions to



1



2



3



4



5





6

cryptography, algebraic graph theory, and the popularisation of mathematical sciences. In 2006, he was awarded a Humboldt Research Award from Alexander von Humboldt Foundation, as well as a three-month fellowship to the Institute of Advanced Study, Durham University (UK).

6 Dr Sean M. Grimmond

Institute for Molecular Bioscience
The University of Queensland

Dr Sean Grimmond pioneered microarray technology at the Medical Research Council Genetics Unit (Harwell, UK) before returning to Australia in 2000 under the C.J. Martin Career Development Award. He is a member of the Editorial Review Board of *Differentiation*, is active in the FANTOM and US National Institutes of Health Stem Cell Anatomy networks, and was Eppendorf Young Australian Scientist 2004. Sean heads IMB's microarray facility.

7 Dr Markus Hegland

Centre for Mathematics and its Applications
Mathematical Sciences Institute
The Australian National University

Dr Markus Hegland is coordinator of Advanced Computation and Modelling at ANU's Centre for Mathematics and its Applications. He is also a member of the Statistical Machine Learning Group of National ICT Australia (NICTA), an Associate Editor of *Computational Statistics and Data Analysis*, and a member of the board of the ANZIAM computational mathematics group. Markus is widely recognised for his expertise in numerical analysis, particularly algorithms and approximation of high-dimensional functions with sparse grids, applications of high-dimensional techniques in machine learning, and computational molecular biology.

8 Prof. Geoffrey J. McLachlan

Professor and Head, Mathematics
The University of Queensland

Prof. Geoff McLachlan is internationally known for his research in statistics applied to the fields of classification, cluster and discriminant analysis, image analysis, intelligent systems, machine learning, neural networks, pattern recognition, and statistical inference.



7



8



9



10

The focus in the latter has been on the theory and applications of finite mixture models, and on estimation *via* the EM (expectation-maximisation) algorithm. A common theme has been statistical computation, with particular attention to computational aspects of statistical methodology. This computational theme extends to the field of data mining. More recently, Geoff has been actively involved in bioinformatics, focusing on the statistical analysis of microarray gene expression data. A joint Wiley monograph on the analysis of microarray data was published in 2004. In 2006 he was awarded a Professorial Research Fellowship for the period 2007-2011 by the Australian Research Council.

9 Dr Ralf M. Muhlberger

School of Information Technology & Electrical Engineering
The University of Queensland

Dr Ralf Muhlberger wrote IBM's Red Book on workflow and image library integration, and is a data integration consultant for the Australian, New Zealand and Queensland governments and for Boeing Corporation. He is a Chief Investigator of the ACID Virtual Communities project, and lectures on Interaction Design and Multimedia in the School of Information Technology and Electrical Engineering. Until 2000 Ralf was Executive Officer and Project Manager within DSTC Pty Ltd (Distributed Systems Technology CRC).

10 Prof. Bernard A. Pailthorpe

School of Physical Sciences
The University of Queensland; and
CEO, Queensland Cyber Infrastructure Foundation

Prof. Bernard Pailthorpe has since 2003 been CEO of Queensland Cyber Infrastructure Foundation (formerly QPSF), the state partner within the Australian Partnership in Advanced Computing (APAC). QCIF and APAC provide computing and data support to ACB. Until 2000 he was Director for Visualization at the San Diego Supercomputer Centre at UC San Diego, and Associate Director of the US National Science Foundation's 50-university NPACI partnership. Bernard is building computational and data infrastructure in support of Queensland researchers, and participates

extensively in state and national policy initiatives and funding programs. His research interests include statistical mechanics, data and image processing, and scientific visualisation.

11 Dr Rohan D. Teasdale

Institute for Molecular Bioscience
The University of Queensland

Dr Rohan Teasdale holds an RD Wright Career Development Award. He leads a multidisciplinary research group at IMB that applies both cellular and computational techniques to investigate how subcellular compartments are generated and how proteins are trafficked within the mammalian cell. Rohan is active in the FANTOM mouse transcriptome annotation consortium. He played a key role in the formation of Bioinformatics Australia and is a member of its steering committee.

12 Prof. Xiaofang Zhou

School of Information Technology & Electrical Engineering
The University of Queensland

Prof. Xiaofang Zhou is Professor of Computer Science at UQ, Research Director of the ARC Research Network in Enterprise Information Infrastructure (EII), and Chair of the Steering Committee of the Asia-Pacific WEB Conferences. Until 1999 he was Principal Researcher and Head of Spatial Information Systems within the Commonwealth Scientific and Industrial Research Organisation (CSIRO). Xiaofang is author of almost 100 publications in spatial databases and information systems, Grid computing, and bioinformatics.

Partner Investigators

13 Dr Prasad Kodali

IBM Healthcare & Life Sciences

Dr Prasad Kodali is Americas Sales Leader for IBM Software Solutions for life sciences. He was previously Manager of the Discovery Solutions development team within IBM Life Sciences, focussed on developing and delivering information solutions to R&D communities. Prasad received his PhD in computational chemistry from The Pennsylvania State University, and an MSc in organic chemistry from Indian Institute of Technology, Bombay. He has held a variety of technical and management positions spanning research, technical development and product management. His experience ranges from creating domain-specific solutions, to providing data integration solutions to life science informatics companies. His primary areas of interest are life science informatics, molecular dynamics simulations, computer-aided drug discovery, and data integration.

14 Dr Isidore Rigoutsos

IBM Thomas J. Watson Research Center

Dr Isidore Rigoutsos manages the Bioinformatics and Pattern Discovery group at IBM's Thomas J Watson Research Center (Yorktown Heights, NY USA), and has been a visiting lecturer at MIT since 2000, where he teaches a Spring semester course and a Summer professional class in bioinformatics. Isidore is a Fulbright Scholar, and in 2003 was elected a Fellow of the American Institute for Medical and Biological Engineering. He serves on editorial boards of several journals in bioinformatics, and is a member of the Advisory Board for the Master's program in bioinformatics of Oxford University. He recently edited a two-volume book on Systems Biology (Oxford University Press, September 2006). His research activities currently focus on the computational study of RNA interference (RNAi) and its application to regenerative medicine and therapy, the study of cancer as a disruption of normal regulatory processes, the analysis of "junk" DNA in animal and plant genomes, and meta-phylogenomics.

11



12



13



14





1



2



3

Special advisor

1 Prof. Gene Myers

Janelia Farms Research Campus
Howard Hughes Medical Institutes
Asburn, Virginia, USA

Dr Gene Myers is one of the foundation group leaders at the new Janelia Farms Research Campus of the Howard Hughes Medical Institutes, near Washington DC. Gene has made fundamental contributions to computational molecular biology and bioinformatics, notably the BLAST algorithm (1990) and whole-genome assembly (1999). He designed and led computer-based assembly of large genome sequences including those of fruit fly (2000), human (2001) and mosquito (2002). His numerous awards include the ACM Kannellakis Theory and Practice Award (2002), membership in the US National Academy of Engineering (2003), the Max Planck Research Prize for International Cooperation (2004), induction into Leopoldina, the German National Academy (2006), and most recently an honorary doctorate from ETH Zürich. Gene has been an Honorary Professor in the Institute for Molecular Bioscience, UQ since 2003.

Strategic partnerships

ACB has formed a close research partnership with the following:

2 Prof. Michael Langston

Department of Computer Science
University of Tennessee, USA

Prof. Mike Langston's research interests include computational biology, graph theory, optimisation and high-performance computation. He is currently Professor of Computer Science at the University of Tennessee, and Collaborating Scientist in the Life Sciences Division at Oak Ridge National Laboratory. Mike is lead investigator on a variety of research projects focused on the design, implementation and enhancement of combinatorial algorithms used to process high-throughput biological data. His collaborations extend to research groups in Australia, Canada, Europe and the Middle East as well as across the USA. He has worked closely

for over twenty years with Professor Mike Fellows. Mike Langston has authored over 200 refereed journal articles, conference papers, book chapters and other reports. His work has been funded by the National Science Foundation, the US Department of Defence, the US National Institutes of Health, the US Department of Energy, the Australian Research Council, and the European Commission.

3 A/Prof. Pablo Moscato

Newcastle Bioinformatics Initiative, and
Priority Research Centre for
Bioinformatics, Biomarker Discovery
and Information-Based Medicine
Department of Computer Science
The University of Newcastle

A/Prof. Pablo Moscato is founding director of the Newcastle Bioinformatics Initiative, a strategic partner of ACB since 2004. He has worked in evolutionary computation since 1988, and in metaheuristic methods for combinatorial optimisation since 1985. A former member of the Caltech Concurrent Computation Program, Pablo introduced memetic algorithms in 1989 and has successfully applied these hybrid methodologies to large-scale optimisation problems. Before coming to Australia, A/Prof. Moscato held academic and research positions in Argentina, Brazil, and the USA. In 2006 he was named Associate Director of the new Priority Research Centre for Bioinformatics, Biomarker Discovery and Information-Based Medicine at The University of Newcastle. He is a member of the editorial board of *Journal of Mathematical Modelling and Algorithms*, has served on the program committee of numerous international conferences in heuristics and optimisation.

Information infrastructure

Mr Matthew Bryant
System Administrator, High-performance computing

Mr Oliver Cairncross
Project Manager, Visible Cell™

Dr Lynn Fink
Research Website Administrator

Mr Igor Kromin (to June 2006)
Application Developer

Ms Mhairi Marshall
Grid & Database Developer

Mr Tim McComb (from January 2007)
Application Developer

Mr Ken Steube (to May 2006)
Web Services Developer

Mr Timothy Sullivan
Application Developer

Mr David Wood
Database Administrator/Developer

Centre Manager

Ms Lanna Wong, MBA

Postdoctoral Fellows / Research Officers

Dr Jiyuan An

Dr Richard Bean

Dr Robert Beiko

Dr Aaron Darling

Dr Lynn Fink

Dr Alistair Forrest

Dr Martin Frith (jointly with RIKEN Genome
Sciences Center, Yokohama)

Dr Nicholas Hamilton

Dr Jim Hanan

Dr Liat Jones

Dr Karin Kassahn

Dr Pritha Mahata

Dr Elena Prieto

Dr Fran Rosamond

Dr Lucia Santoso

Dr Tianhai Tian

Dr James Watson

Dr Simon Wong

Dr Zheng Yuan

Research Assistants

Mr Daniel Bradley

Ms Melissa Brown

Mr Geoffrey Faulkner

Ms Kelly Hanson (to September 2006)

Mr Radosav Pantelic (to May 2006)

Mrs Chikako Ragan

Dr Melvena Teasdale

Mr Tom Whittington

Mr Shane Zhang (to February 2006)

Masters/PhD students

Supervision/Associate supervision (principal advisor in ACB):

Farah Abdullah (due to complete in 2007)
Advisor: Prof. Kevin Burrage

Rajith Aturaliya (due to complete in 2007)
Advisor: Dr Rohan Teasdale

Denis Bauer (due to complete in 2010)
Advisor: Dr Timothy Bailey

Michael Bode (due to complete in 2007)
Advisor: Prof. Kevin Burrage

Alhadi Bustamum (due to complete in 2008)
Advisor: Prof. Kevin Burrage

Cheong Xin Chan (due to complete in 2007)
Advisor: Prof. Mark Ragan
Associate Advisor: Dr Robert Beiko

Yan Chen (due to complete in 2008)
Advisor: A/Prof. Phoebe Chen

Mikolaj Cieslak (due to complete in 2010)
Advisor: Dr Jim Hanan

Wagner Emanuel Costa
(due to complete in 2007)
Advisor: A/Prof. Pablo Moscato
Associate Advisor: Dr Regina Berretta

Melissa Davis (PhD thesis submitted, 2006)
Advisor: Dr Rohan Teasdale

Geoffrey Faulkner (due to complete in 2009)
Advisor: Dr Sean Grimmond

Al Forrest (PhD awarded, 2006)
Advisor: Dr Sean Grimmond

Alex Garcia (PhD thesis submitted, 2006)
Advisor: Prof. Mark Ragan
Associate Advisor: Dr Lindsay Hood

Liqi Han (due to complete in 2010)
Advisor: Dr Jim Hanan

Michael Höhl (PhD awarded, 2006)
Advisor: Prof. Mark Ragan
Associate Advisor: Dr Peter Adams

Mou'ath Adeib Hourani
(due to complete in 2007)
Advisor: A/Prof. Pablo Moscato

Helen Zi Huang (due to complete in 2007)
Advisor: A/Prof. Xiaofang Zhou
Associate Advisor: Prof. Peter Bruza
(Queensland University of Technology)

Keith Knapp (due to complete in 2009)
Advisor: A/Prof. Phoebe Chen



Chin Foon Khoo (due to complete in 2009)
Advisor: Dr Markus Hegland

Mario Inostroza-Ponta
(due to complete in 2008)
Advisor: A/Prof. Pablo Moscato
Associate Advisor: Dr Regina Berretta

Dennis Lee (PhD thesis submitted, 2006)
Advisor: Dr Ralf Muhlberger
Associate Advisors: Dr Matthew Simpson
and Dr Mark Brown

Stephen Livingstone
(due to complete in 2008)
Advisor: Dr Ralf Muhlberger
Associate Advisor: Dr Andrew Brown

Andrew Loch (due to complete in 2007)
Advisor: Dr Ralf Muhlberger
Associate Advisor: Dr Stephen Viller

Shev MacNamara (due to complete in 2007)
Advisor: Prof. Kevin Burrage

Katrina Monico (due to complete in 2007)
Advisor: Prof. Geoff McLachlan

Ann Morrison (due to complete in 2008)
Advisor: Dr Ralf Muhlberger
Associate Advisor: Dr Peta Mitchell

Tatiana Semenova
(PhD thesis submitted, 2006)
Advisor: Dr Markus Hegland

Chang Jin Shin (due to complete in 2009)
Advisor: Prof. Mark Ragan
Associate Advisor: Dr Simon Wong

Josefine Sprenger (due to complete in 2009)
Advisor: Dr Rohan Teasdale

Chinnu Subramaniam
(due to complete in 2012)
Advisor: A/Prof. Phoebe Chen

Jason Yang (due to complete in 2008)
Advisor: Dr Ralf Muhlberger
Associate Advisor: Dr Stephen Viller

Justin Xi Zhu (due to complete in 2007)
Advisor: Prof. Geoff McLachlan
Associate Advisor: Dr Angus Ng

**Co-supervision (principal advisor not an
ACB investigator)**

Agnes Boskovitz (due to complete in 2007)
Associate Advisor: Dr Markus Hegland
Principal Advisor: Dr Rajeev Goré

Fabian Buske (due to complete in 2007)
Associate Advisor: Dr Timothy Bailey
Principal Advisor: Dr Mikael Bodén

Simon Carter (due to complete in 2008)
Associate Advisor: Dr André Leier
Principal Advisor: Dr Paul Strooper

Elizabeth Dun (due to complete in 2007)
Associate Advisor: Dr Jim Hanan
Principal Advisor: Dr Christine Beveridge

Nicholas Geard (PhD awarded, 2006)
Associate Advisor: Dr Mikael Bodén
Principal Advisor: Prof. Janet Wiles

Tony Gill (due to complete in 2009)
Associate Advisor: Prof. Bernard Pailthorpe
Principal Advisor: A/Prof. Stuart Phinn

Pol Haji (PhD awarded, 2006)
Associate Advisor: Dr Markus Hegland
Principal Advisor: Dr Robert S. Anderssen

John Hawkins (due to complete in 2008)
Associate Advisors: Dr Rohan Teasdale and
Prof. Janet Wiles
Principal Advisor: Dr Mikael Bodén

Stefan Maetschke (due to complete in 2007)
Associate Advisors: Prof. Geoff McLachlan
and Dr Mikael Bodén
Principal Advisor: Dr Marcus Gallagher

Luke Mathieson (due to complete in 2007)
Associate Advisor: A/Prof. Pablo Moscato
Principal Advisor: Dr Ljiljana Brankovic

Duncan Mortimer (due to complete in 2008)
Associate Advisor: Prof. Kevin Burrage
Principal Advisor: Dr Geoffrey Goodhill

Andrew Noske (due to complete in 2008)
Associate Advisors: Prof. Mark Ragan and
Prof. Kevin Burrage
Principal Advisor: Dr Brad Marsh

Radosav Pantelic (due to complete in 2008)
Associate Advisor: Dr Nicholas Hamilton
Principal Advisor: Dr Ben Hankamer

Øystein Petterson (M.IT awarded in 2006)
Associate Advisor: Prof. Bernard Pailthorpe
Advisor: Dr Nicole Bordes

Blaize Rhodes (due to complete in 2008)
Associate Advisor: Dr Ralf Muhlberger
Principal Advisor: Prof. Simon Kaplan (QUT)

Stefan Stanley (due to complete in 2007)
Associate Advisor: Prof. Mark Ragan
Principal Advisor Prof. John Mattick

Henk Stolk (PhD awarded, 2006)
Associate Advisor: Dr Jim Hanan
Principal Advisor: Dr Kevin Gates

Stuart Stephen (due to complete in 2007)
Associate Advisor: Prof. Mark Ragan
Principal Advisor: Prof. John Mattick

James Watson (PhD awarded, 2006)
Associate Advisors: Dr Jim Hanan and
Dr Mikael Bodén
Principal Advisor: Prof. Janet Wiles

Kai Willadsen (PhD awarded, 2006)
Associate Advisor: Dr Mikael Bodén
Principal Advisor: Prof. Janet Wiles

Jemma Wu (PhD awarded, 2006)
Associate Advisor: Dr Markus Hegland
Principal Advisor: Dr John Lloyd

Honours students

Rachel Blakers
Numerical solution of the Fokker-Planck
Equations
Advisors: Dr Markus Hegland and
Dr Stephen Williams

Mitchell Stanton-Cook
Location and orientation of ligand molecules
by paramagnetic NMR techniques for
application in pharmaceutical screening
Advisors: Dr Nicholas Hamilton and
Dr Thomas Huber

Adam Fairley
Molecular dynamics simulations of the self-
assembly of the Lac21 and Lac28 peptides
for bio-nano applications
Advisors: Prof. Bernard Pailthorpe and
Dr Anette Dexter

Robert McLeay
Bioinformatics discovery of novel
endosomal proteins
Advisor: Dr Rohan Teasdale

Arthur Ngui
Enhancing protein secondary structure
prediction
Advisors: Dr Mikael Bodén and
Dr Timothy Bailey

Emma Redhead
Discriminative motif discovery
Advisor: Dr Timothy Bailey

Jarad Sanderson
Data capture and the management of data
streaming from a network of sensors
Advisors: Dr Nicole Bordes and
Prof. Bernard Pailthorpe

Mark Wakabayashi
A biologically plausible model of *C. elegans*
locomotion and motor control
Advisor: Prof. Janet Wiles

Jack Wang
Quantitative analysis of macropinosomes
Advisors: Dr Rohan Teasdale and
Dr Nicholas Hamilton

International & national interns

International

Jonathan Chevelu
Université de Rennes (France)
Advisor: Prof. Geoff McLachlan

Zuzana Cienikova
Institut National des Sciences Appliquées
de Lyon (France)
Advisor: Dr Timothy Bailey

Andrea Heidenreich
University of Applied Sciences,
Weihenstephan, Freising (Germany)
Advisors: Mr Oliver Cairncross and
Prof. Mark Ragan

Fritjof Heyde
University of Applied Sciences,
Weihenstephan, Freising (Germany)
Advisor: Dr Jim Hanan

Rolv Seehuus
Norwegian University of Science and
Technology (Norway)
Advisor: Dr Timothy Bailey

Triinu Tasa
Chalmers University (Sweden)
Advisor: A/Prof. Pablo Moscato

National

Robert McLeay
The University of Queensland
Advisor: Dr Jim Hanan

Liang Ma
University of New South Wales
Advisor: Dr Timothy Bailey

Dave Tang
The University of Queensland
Advisor: Dr Robert Beiko

Benjamin Woodcroft
The University of Queensland
Advisors: Dr Nicholas Hamilton and
Dr Thomas Huber

Exchange student

Leo Huber
ETH Zürich (Switzerland)
Australian National University
Advisor: Dr Markus Hegland

Research trainees and volunteers

Josefine Sprenger
Occupational Trainee
Evaluations of mammalian subcellular
localisation predictors
Advisor: Dr Rohan Teasdale

Chikako Ragan
Volunteer
Computational prediction of miRNA targets
in human and mouse genomes
Advisor: Prof. Mark Ragan

Visiting scholars, associates

Visiting Scholars

Dr Regina Berretta, Newcastle
Bioinformatics Initiative, School of Electrical
Engineering and Computer Science,
The University of Newcastle (Australia)

Prof. Mark Borodovsky, Founding Director,
Bioinformatics Program, Georgia Institute of
Technology (USA)

Dr Aaron Darling, University of Wisconsin
(USA)

Prof. John Dennis, Noah Harding Professor
Emeritus and Research Professor,
Department of Computational & Applied
Mathematics, Rice University (USA)

Mr Danny Hermelin, University of Haifa
(Israel)

Mr Hong Liang Hiew, Centre for
Comparative Genomics, Murdoch University
(Australia)

Prof. David Lambert FRSNZ, Allen Wilson
Centre for Molecular Evolution, Massey
University (NZ)

Dr Simon Malham, School of Mathematical
and Computer Sciences, The University of
Edinburgh (UK)

A/Prof. Pablo Moscato, Newcastle
Bioinformatics Initiative, School of Electrical
Engineering and Computer Science,
The University of Newcastle (Australia)

Prof. Osamu Maruyama, Department of
Mathematical Sciences, Kyushu University
(Japan)

Mr Moritz Mueller, University of Freiburg
(Germany)

Mr Rolv Seehuus, Norwegian University of
Science and Technology (Norway)

Dr Wei-Mou Zheng, Institute of Theoretical
Physics, Academia Sinica, Beijing
(PR China)

Associates

Dr Jonathan Arthur, Sesqui Lecturer in
Bioinformatics, University of Sydney

A/Prof. Ian Atkinson, Information
Technology, James Cook University

Prof. Kaye Basford, Head, School of
Land & Food Sciences, The University
of Queensland

Mr Geoff Bulow, Queensland Department of
Primary Industries and Fisheries

Dr Pamela Burrage, Department of
Mathematics, The University of Queensland

Dr Mikael Bodén, School of Information
Technology & Electrical Engineering,
The University of Queensland

Dr Christopher Bouton, Head, Integrative
Data Mining, Pfizer Research Technology
Center, USA

Prof. Vladimir Brusic, Australian Centre for
Plant Functional Genomics, and Institute
for Molecular Bioscience, The University
of Queensland

Prof. Gillian Bushell, Dean, Faculty of
Science, Griffith University

Dr Robert Charlebois, Founder,
NeuroGadgets Inc., Ottawa

Dr Brian Dalrymple, CSIRO
Livestock Industry

Prof. Frank Dehne, School of Computing and
Information Technology, Griffith University
and School of Computer Science, Carleton
University, Ottawa

Prof. Werner Dubitzky, Head, Bioinformatics
and Systems Biology, University of Ulster,
Northern Ireland

Dr David Gilbert, Professor of
Bioinformatics, University of Glasgow, UK

Dr Ben Hankamer, Institute for Molecular
Bioscience, The University of Queensland

Prof. John Hancock, Institute for Molecular
Bioscience, The University of Queensland

Dr David Hansen, e-Health Research
Centre, CSIRO



Dr James Hogan, School of Software Engineering and Data Communications, Queensland University of Technology

Dr Lindsay Hood, Australian Partnership in Advanced Computing, Canberra

Dr Thomas Huber, School of Molecular & Microbial Sciences, The University of Queensland

Dr Philip Hugenholtz, Head, Microbial Ecology Program, US DOE Joint Genome Institute, USA

Prof. Jane Hunter, School of Information Technology & Electrical Engineering, The University of Queensland

Dr Lars Jermiin, School of Biological Sciences, The University of Sydney

Prof. Simon Kaplan, Executive Dean, Faculty of Information Technology, Queensland University of Technology

Dr Jonathan Keith, Faculty of Science, Queensland University of Technology

Prof. Do Han Kim, Department of Life Science, Gwangju Institute of Science and Technology, Korea

Prof. Michael Langston, Department of Computer Science, University of Tennessee, USA

Prof. Sang Yup Lee, Director, Bioinformatics Research Center, KAIST, Korea

Dr Wilfred Li, Research Fellow, San Diego Supercomputer Center, University of California San Diego, USA

Prof. Peter Lindsay, School of Information Technology and Electrical Engineering, The University of Queensland; and Director, ARC Centre in Complex Systems

Dr Anthony Maeder, Research Director, e-Health Research Centre, CSIRO

Prof. Satoru Miyano, Institute of Medical Science, University of Tokyo, Japan

Dr Gary Morgan, Director, e-Health Research Centre, CSIRO

A/Prof. William Stafford Noble, Department of Genome Sciences, University of Washington, USA

Prof. Shoba Ranganathan, Biotechnology Research Institute, Macquarie University

Dr Stephen Robinson, University of Ulster, Northern Ireland

Prof. Allen Rodrigo, Director, Bioinformatics Institute New Zealand, University of Auckland, NZ

Prof. Jenny Stow, Institute for Molecular Bioscience, The University of Queensland

Dr Harukazu Suzuki, RIKEN Genome Sciences Center, Yokohama

Prof. Ah Chung Tsoi, Director, Monash e-Research Centre, Monash University

Dr Anne Trefethen, Executive Director, Oxford e-Research Centre, UK

Dr Mark Wilkinson, Department of Medical Genetics, University of British Columbia and St Paul's Hospital, Vancouver, Canada

Prof. Janet Wiles, School of Information Technology & Electrical Engineering, The University of Queensland

Prof. Michael Zuker, Howard Hughes Medical Institute Janelia Farms Campus, USA

Vision

The Australian Research Council (ARC) Centre in Bioinformatics is a multi-nodal interdisciplinary Centre that uniquely brings together genomic and post-genomic bioinformatics, complex systems research and experimental phenomics to produce a deep understanding of how information in the genome is transformed into structure and function in the mammalian cell.

Mission

The ARC Centre in Bioinformatics is committed to:

- Research excellence providing insights into genome-phenome biology and complex systems.
- An advanced education and training program that fosters excellence, exposes students from diverse disciplines to new perspectives in bioinformatics, and generates human capital.
- Improved access to skills, tools and facilities at the interface among genomic and post-genomic biology, molecular cell biology, mathematics, statistics, and computing and information sciences.
- Critical mass and a national focus for genome-phenome bioinformatics in Australia.

Goals

The ARC Centre in Bioinformatics aims to achieve:

- Integration of insights and new developments in mathematics, statistics, computing and information sciences, genomic and post-genomic biology, and molecular cell biology.
- Mutual enrichment of biological and complex systems research.
- Development and application of innovative laboratory-based experimental phenomic biology based on reverse transfection arrays and high-resolution electron tomography.
- Building an integrated data and visual framework for hypothesis-testing, validation and discovery.

Objectives

The ARC Centre in Bioinformatics will:

- Examine how the genome comes to life in the mammalian cell during differentiation and development.
- Model, visualise and experimentally validate selected complex cellular systems and regulatory networks that control transformation of genomic information into biological structure and function.
- Develop novel approaches and tools of broad utility in molecular and cellular biotechnologies.
- Build critical mass and national focus in bioinformatics to generate the human capital and intellectual property that will be vital if Australia is to compete internationally in advanced bioscience and biotechnology.

Capabilities — core technologies and infrastructure

The integrated research programs of the ARC Centre in Bioinformatics are underpinned by established strengths in mathematics, computing science, information technologies, and molecular and genomic biosciences.

Key capabilities include:

- Algorithmics and computational complexity
- Statistics, including expression microarray statistics
- Pattern discovery, machine learning and adaptive optimisation
- Advanced database architectures, including distributed and visual databases
- Specialised mark-up languages, ontologies and semantic web technologies
- Computational modelling, including stochastic, deterministic, and agent-based approaches
- Computational simulation using high-performance parallel and distributed architectures
- Automated discovery pipelines and workflows in bioinformatics
- Advanced scientific visualisation
- High-throughput microarray-based genomic technologies, including reverse transfection arrays
- Remote interaction *via* AccessGrid



The Centre's research is delivered through programs in experimental phenomics, modelling of cellular networks, and data integration and visualisation. Each of these programs is organised and managed as a set of projects. In the following pages we describe the aims of the 13 projects we managed during 2006, identify their personnel, and report their achievements during 2006. In a separate (non-public) document we present detailed activity plans for each project for 2007.

This coming year – the last under our initial funding from ARC – will bring more than the usual amount of project realignment. Several projects had been scoped to conclude in 2007, and will do so. Other projects, or their infrastructural components, will be transferred into the new Queensland Facility for Advanced Bioinformatics (QFAB).

In **Program 1** we apply high-throughput experimental methods to develop and test key hypotheses about the control of developmental processes in mammalian cells and generate data required to model cellular and molecular-interaction networks. We use GFP and epitope tagging, high-throughput reporter screens, microarray technologies, confocal and real-time video microscopy, and computational pipelines to obtain genome-wide views of the production and subcellular localisation of key regulatory protein families in human and mouse. Through collaboration involving the Nanostructural Analysis Network Organisation (NANO) Major National Research Facility (MNRF) we also have unique access to high-resolution 3D electron tomography data on mammalian cells. The activity in this program provides the gene and protein networks for construction of computational models, and allows us to validate their predictions against observed phenotype. Phenomic data also underlie our computational models of complex cellular structures, including dynamic spatial relationships among vesicles, lipid rafts and tubules during endocytosis.

The objective of **Program 2** is to model and simulate selected networks of gene regulation and protein interactions in mammalian cells, using powerful analytical tools of computational mathematics and complex systems research. To achieve this we develop efficient algorithms, apply advanced methods from statistics and machine learning, integrate expertise in network representation and analysis, identify orthologs, and use high-performance computing. During 2006 it became increasingly difficult to demarcate several of the projects in Program 2 from projects in Programs 1 and 3, as our investigators and research staff collaborated to share datasets, models and software. Notably former Project 2.5, which emphasised the development of computational models and stochastic simulators, was merged with Project 3.2.2, which emphasised the parameterisation of these models with empirical data and their application, within an integrated data and visualisation framework, to biological questions.

Program 3 has begun to deliver an innovative prototype data, modelling and visualisation environment, the Visible Cell™, that will seamlessly embed molecular structures, networks and data from quantitative simulations into a virtual 3D mammalian cell reconstructed from high-resolution tomograms. Understanding how the molecular and regulatory networks in mammalian cells interact in space and time to yield complex phenotype requires us to synthesise large quantities of diverse, multi-scale data including images, molecular structures and mathematical simulations. In other areas of science similarly characterised by very large, complex, multi-scale data (e.g. atmospheric science, physical oceanography, neurobiology), integrative analyses are all but impossible without advanced visualisation. Genome-phenome biology presents an even greater challenge, as our data are semantically rich and highly interrelated. By directly utilising the most highly evolved pattern-recognition system on Earth – the human visual cortex – our Visible Cell™ will facilitate hypothesis-formulation and hypothesis-testing with the large, complex, semantically rich datasets of genomic and post-genomic

bioinformatics. By providing a common implementation platform, this project will unify our Centre at the practical, hands-on level. We also appreciate the challenges of making our data, models and visualisations discoverable, searchable and readily integrated across the nascent Grid, and in 2006 we built on ongoing collaborations, and initiated new ones, designed to bring forefront Grid initiatives to bioinformatics and molecular bioscience.

It is not easy to deliver a deeply integrative research program across disciplines as methodologically and sociologically diverse as mathematics, information technology and genomics. In 2006 we continued to focus our limited resources on projects that build toward integrative research outcomes, supporting collaborations, student projects, shared data infrastructure including QFAB, use of common standards and protocols, and our Visible Cell™ project.



Program 1 – Experimental phenomics

1.1 Computational and microarray-based analysis of transcriptional complexity of mammalian cells

Investigator:

Dr Sean Grimmond
(The University of Queensland)

In collaboration with:

Prof. David Hume
(The University of Queensland)

Researchers:**Research Associate:**

Dr Al Forrest
(The University of Queensland)

Research Assistants:

Ms Melissa Brown
(The University of Queensland)

Mr Geoffrey Faulkner
(The University of Queensland)

Aims:

- (a) Functionally annotate key classes of gene products in the mouse, particularly phosphoregulators and transcription factors. Many of their transcripts potentially encode novel dominant negative and other cis-regulating gene products that are likely to play major roles in mammalian development, differentiation and disease.
- (b) Globally monitor the expression of all transcripts from each phosphoregulator locus and define cell-specific variants.
- (c) Globally monitor the activity of the non-coding transcripts and alternative promoter usage for every phosphoregulator locus monitor across a panel of cell lines and in cell differentiation models.

Achievements (2006):

- (a) We have completed a detailed computational analysis of transcriptional complexity and a combination of in-house splice arrays, MPSS, all-exon arrays and RT-PCR to experimentally support the biological significance of many variant transcripts.
- (b) We have examined phosphoregulator expression profiles and CAGE data to identify examples of alternative promoter usage. These data have been compiled with the transcriptome surveys described above to derive independent evidence of alternative promoter usage. Furthermore we began to address the potential of non-coding antisense and repeat containing transcripts to transcripts to regulate transcriptional output from mammalian loci.
- (c) For well-supported potential dominant negative variants, we commenced experiments to evaluate the functional consequences (the ability to alter cell signalling, regulate cell division) of variants from a panel of loci. The results have highlighted novel levels of control of the phosphoregulatory network by transcriptional complexity.



1.2 Computational cellular biology: combining computational prediction and experimental validation

Investigator:

Dr Rohan Teasdale
(The University of Queensland)

In collaboration with:

Prof. P. Carninci
(RIKEN Genomic Sciences Center, Japan)

Prof. Y. Hayashizaki
(RIKEN Genomic Sciences Center, Japan)

Dr H. Suzuki
(RIKEN Genomic Sciences Center, Japan)

Researchers:

Research Associates:

Dr Lynn Fink
(The University of Queensland)

Ms Seetha Karunaratne
(The University of Queensland)

Dr Zheng Yuan
(The University of Queensland)

PhD students:

Mr Raj Aturaliya
(The University of Queensland)

Ms Melissa Davis
(The University of Queensland)

Research Assistant:

Mr Shane Zhang
(The University of Queensland)

Honours student:

Mr Robert McLeay
(The University of Queensland)

Aims:

- (a) Establish a publicly accessible database for data on subcellular localisation of proteins.
- (b) Refine the computational prediction pipelines for membrane organisation.
- (c) Define the set of membrane and secreted proteins within the murine transcriptome.
- (d) Apply pattern discovery and other informatic methods to recognise and systematise our understanding of intracellular targeting signals, and to validate predicted subcellular localisations.
- (e) We compared different methods for predicting the subcellular localisation of mammalian proteins, concluded that none of the current methods is suitable for annotation of proteins of unknown subcellular localisation.
- (f) We determined the subcellular localisation of proteins in the transcriptional regulator set generated by RIKEN GSC (>1500 proteins). Collectively, we have now localised representative proteins from >10% of all the protein-coding genes as defined by FANTOM3.

Achievements (2006):

- (a) Developments to the LOCATE subcellular localisation database including:
 - Incorporation of a human protein isoform dataset;
 - Allowing third-party submissions of literature subcellular localisation data;
 - Allowing intranet submission of both image and literature subcellular localisation data;
 - Continued population of the literature data within LOCATE; and
 - Incorporation of subcellular localisation prediction data based on third-party prediction methods.
- (b) We described the LOCATE database in the *Nucleic Acids Research Annual Database* issue. LOCATE was featured in numerous bioinformatics resources including *Science* magazine's Netwatch.
- (c) We published the high-throughput subcellular localisation platform in *Traffic*. This work reports the subcellular localization of 24% of the putative mouse type II membrane proteins.
- (d) We identified protein isoforms that are targeted to multiple subcellular locations, and showed that differential targeting represents one of the major functional consequences of transcript variation within the mouse transcriptome.

1.3 Generation of the experimental data for the modelling of cellular compartments and the development of organelle biogenesis phenotype screens

Investigators:

Dr Rohan Teasdale
(The University of Queensland)

Prof. Kevin Burrage
(The University of Queensland)

In collaboration with:

Prof. Jenny Stow
(The University of Queensland)

Researchers:

Research Associates:

Dr Lynn Fink
(The University of Queensland)

Dr Nicholas Hamilton
(The University of Queensland)

Research Assistants:

Ms Kelly Hanson
(The University of Queensland)

Mr Radosav Pantelic
(The University of Queensland)

Honours student:

Mr Jack Wang
(The University of Queensland)

Aims:

This project is focused on the systematic measurement of organelle compartments made within fixed cells using confocal microscopy combined with 3D reconstructions, and in live cells using real-time microscopy. The reagents generated will be suitable for the development of phenotype screens that could be considered for future application on a broad scale.

- (a) Develop a suite of reagents to mark individual membrane subcellular compartments and organelles.
- (b) Express organelle markers within a range of cell types.
- (c) Generate stable cell lines expressing GFP-tagged organelle markers.
- (d) Establish automatic image analysis and capture technologies (with Dr Nick Hamilton).
- (e) Generate custom data in response to requirements of cellular modellers (Dr Nick Hamilton).

Achievements (2006):

- (a) We developed image training data for endogenously expressed proteins or organelle vital stains, and for transiently transfected proteins. This involved capturing >50 images of each of the 13 distinct regions of the mammalian cell.
- (b) We prepared plasmids for the generation of stable cell lines expressing GFP-tagged organelle markers, and generated stable cell lines using NIH3T3.
- (c) We developed and published the Automated Subcellular Phenotype Classification system (ASPiC) software to classify the 13 distinct regions of the cell. This software is available on request. We implemented and used ASPiC for image classification and curation within the LOCATE database, as well as to quantify localisation changes under co-expression. We further extended ASPiC to classify up to 10 subcellular localisations with an accuracy of up to 98%, leading the field in automated classification.
- (d) We developed new fast image statistics for ASPiC that give comparable or better accuracy than previous statistics, while being an order of magnitude faster to calculate. This greatly extending the range of potential applications.

- (e) Based on some of the ideas underlying ASPiC, we developed ImageBlast, a prototype image searching system. ImageBlast aims to make cell image databases searchable directly by content of the images, rather than *via* human-supplied metadata, hence delivering more-refined querying and unbiased searching, as well as the potential to compare across databases in high throughput. Preliminary results show that image statistics developed in ASPiC can be used to rank matches from an image set to a given image, and that this ranking corresponds well with known subcellular localisations of the images. A review of the challenges of high-throughput cell imaging and making image databases searchable was published in *Australian Biochemist*.
- (f) We made publicly available comprehensive information on organelle marker proteins and their application to a range of cell types *via* the LOCATE database. We agreed to supply these images to Dr Leanne Bischoff, CSIRO Biotechnology Imaging.



Program 2 – Modelling cellular networks

2.1 Hybrid algorithms for large-scale network and graph optimisation problems in bioinformatics

Investigators:

Prof. Michael Fellows
(The University of Newcastle)

A/Prof. Pablo Moscato
(The University of Newcastle)

In collaboration with:

Dr Timothy Bailey
(The University of Queensland)

Dr Regina Berretta
(The University of Newcastle)

A/Prof. Phoebe Chen
(Deakin University)

Prof. Frank Dehne
(The University of Ottawa)

Dr Markus Hegland
(The Australian National University)

Prof. Mike Langston
(University of Tennessee and ORNL)

Prof. Mark Ragan
(The University of Queensland)

Prof. Rodney Scott
(The University of Newcastle)

Researchers:

Research Associates:

Dr Pritha Mahata
(The University of Newcastle)

Dr Elena Prieto
(The University of Newcastle)

Dr Frances Rosamond
(The University of Newcastle)

PhD students:

Mr Mario Instroza-Ponta
(The University of Newcastle)

Mr Peter Shaw
(The University of Newcastle)

Aims:

- (a) Identify a common core of generic optimisation techniques that are highly “reusable” or “successful” among the different problem areas of interest in the ARC Centre, and implement the algorithmic solutions in a modular framework that will enable parallel, distributed and grid computing.
 - (b) Develop a systematic approach towards the hybridisation of fixed-parameter tractable (FPT) algorithmics, other stochastic search heuristics and metaheuristics, and approximative algorithmics.
 - (c) Address open research issues of interest to the entire community of the ARC Centre for Bioinformatics, in particular by collaborating in the area of algorithm design, analysis and implementation of large-scale graph and network optimisation problems.
 - (d) Develop novel synergistic hybrid algorithms that benefit from local search and stochastic algorithms (*e.g.* memetic algorithms) for which no fixed-parameter algorithm is known.
- (b) We continued development of FPT and hybrid algorithms to analyse expression microarray data, in particular to understand genetic networks, emphasising algorithms that are robust to the presence of experimental noise. We investigated the usefulness of the current reduction rules for the Max Cut problem. Although Max Cut has theoretical importance *per se*, we found it to be of less value as part of a strategy for unsupervised clustering. We showed that a new formalism, Arithmetic/Harmonic Cut, yields superior results in several datasets and used it to generate a molecular taxonomy of breast cancer tumours. We studied the computational complexity and approximability of Arithmetic/Harmonic Cut.
 - (c) We investigated several graph-optimisation problems related to node/edge deletion as applied to clustering, with the goal of establishing new reduction rules leading to FPT results or, alternatively, proving parameterised intractability. This has application across bioinformatics, and is central to specific applications in analysis of microarray data. Collaboration of Prof. Mike Fellows with Prof. Mike Langston and Dr Frances Rosamond led to a major breakthrough in polynomial-time pre-processing for the Feedback Vertex Set (FVS) problem. Whether FVS admits a pre-processing kernel of size bounded by a polynomial in the feedback set was considered a major open problem in FPT research. The above team, joined by Prof. Kevin Burrage and PhD student Shev MacNamara in regard to the application of FVS to the analysis of genetic regulatory networks, gave such an algorithm. Their work, presented at IWPEC 2006, was considered one of the most important developments of 2006 in parameterized algorithmics. Their methodologically novel approach, exploiting a connection between FPT kernelisation and P-time approximation, has since been widely adopted in the field.

Achievements (2006):

- (a) In 2006 we continued developing collaborations in which datasets from other ACB projects, or external sources, are investigated using algorithms for information extraction and visualisation developed within the Newcastle Bioinformatics Institute. The NBI and Deakin nodes (A/Prof. Phoebe Chen) analysed an Alzheimer’s disease microarray dataset to evaluate the complementarity of two approaches for identifying differential gene expression. Our joint analysis revealed that the method of *Emerging Patterns* presents some problems in handling intrinsic differential expression of in different regions of the brain used as control, and the same was seen in analysis of a rodent model of Parkinson’s disease. Thus *Emerging Patterns* seems to be beneficial in cases where we expect *a priori* a strong similarity among control samples, *e.g.* in comparison of different types of solid tumours against normal tissue samples of the same patients.

- (d) The team of Fellows, Langston and Rosamond, together with PhD student Mr Peter Shaw of Newcastle, achieved a linear kernelisation for the important Cluster Editing problem based on “crown reduction” techniques developed at Newcastle since 2003. This advance has been implemented in software by collaborators Prof. Mike Langston and Prof. Frank Dehne, and applied to clustering problems for a number of very large datasets. Efforts to achieve similar breakthroughs for related clustering problems continue.
- (e) In discussions with Prof. Mark Ragan and Dr Simon Wong we identified a possible application of Clique Cover to the analysis of protein domains.
- (f) FPT methods based on bounded treewidth have strong new applications in areas such as protein folding prediction, based on a new notion of “operational parameters” to allow hybridization of FPT methods and heuristic objectives. Prof. Mike Fellows and Dr Fran Rosamond, in collaboration with researchers in the UK, Israel, Denmark and Germany, achieved two very significant results on “width-based” FPT methods in 2006. They solved a long-standing open problem concerning the complexity of the Clique Width problem; this was presented at STOC 2006, and submitted for journal publication. They also developed a new way to use bounded treewidth to design FPT algorithms *via* introduction of new classes of vertices to the graph model to encode semantic information (*e.g.*, about colouring or geometry, which cannot be expressed directly in MSO logic). One paper using this method was submitted in 2006, and further applications were identified.
- (g) We began to implement anytime-complete memetic algorithms for some of the most intensively used codes developed by the Newcastle group, using a new six-processor cluster. In early 2006, NBI bought 10 floating-point licences for CPLEX, which will allow their concurrent use on this cluster and will support the collaboration with Prof. Mike Langson.
- (h) Collaboration between our Brisbane and Newcastle nodes was enhanced as A/Prof. Pablo Moscato and Dr Regina Berretta spent three months at the Institute for Molecular Bioscience. Collaboration with Prof. Mark Ragan and Dr Rob Beiko focused on application of a new algorithm, developed by A/Prof. Moscato and M.Sc. student Mr Mario Inostroza, to clustering of large protein-sequence datasets. Drs Moscato and Berretta collaborated on the design of a new algorithm for clustering microarray data, and manuscript on identifying molecular profiles in Gleason pattern 3 prostate cancer was initiated. Two members of NBI, Dr Alexandre Mendes and Mr Mario Inostroza, also visited the UQ groups.
- (i) A technical problem with the Illumina Beadstation acquired in late 2005 by Prof. Rodney Scott (Newcastle) delayed the collection of unique Australian datasets on chromosomal breakpoints in cancers. Similarly delayed has been our investigation of SNP data sets derived from formalin-fixed, paraffin-embedded material using Illumina’s DAZL assay. The equipment is expected to be repaired in early 2007. In the meantime, we are developing some of the necessary algorithm methodology based on public-domain data.



2.2 Computational discovery of gene-regulatory motifs and motif clusters

Investigator:

Dr Timothy Bailey
(The University of Queensland)

In collaboration with:

Dr Mikael Bodén
(The University of Queensland)

Prof. John Mattick
(The University of Queensland)

Dr William Stafford Noble
(University of Washington)

Researchers:

Research Associates:

Dr Martin Frith
(The University of Queensland, and RIKEN
Genome Sciences Center)

Dr Zheng Yuan
(The University of Queensland)

PhD student:

Ms Denis Bauer
(The University of Queensland)

Honours student:

Ms Emma Redhead
(The University of Queensland)

Visiting researcher:

Mr Rolv Seehuus
(Norwegian University of Science
and Technology)

Aims:

- (a) Understand genetic regulatory networks through the computational discovery of sequence motifs responsible for transcription of protein-coding genes.
- (b) Design algorithms and use machine-learning techniques, including hidden Markov models, for the discovery of regulatory motifs and clusters of regulatory motifs.
- (c) Develop software and web tools implementing these algorithms and techniques.
- (d) Apply pattern-discovery algorithms and software tools to the annotation of genomic regions and proteins, the characterisation and comparison of prokaryotic and eukaryotic genomes, and the genome-scale study of biological processes including the regulation of genes and gene networks.
- (e) Apply machine learning approaches to predict protein characteristics from sequence data.

Achievements (2006):

- (a) With collaborator Dr William Stafford Noble, we completed development of the Motiph algorithm for detection of transcription factor binding sites in genomic sequence, and implemented it in software.
- (b) With honours student Ms Emma Redhead, we completed development of an algorithm for discovery of discriminative motifs in genomic sequence.
- (c) We implemented enhancements to the MEME motif discovery algorithm, including a new search algorithm.
- (d) Collaboration between Dr Martin Frith and Dr Tim Bailey led to a novel motif-discovery algorithm, GLAM2, that allows gaps in motifs.
- (e) In another collaborative project, Drs Bailey and Noble developed a new algorithm, called Tomtom, for comparison of transcription-factor binding motifs in genomic sequence.
- (f) Collaboration with Dr Mikael Bodén and Dr Zheng Yuan led to the development and implementation of novel algorithms for predicting structural properties of cellular proteins based on primary sequence.

- (g) PhD student Ms Denis Bauer continued investigation of the use of auxiliary information, including protein-protein interaction data, to supplement primary sequence in analysis of transcription factors.
- (h) In collaboration with Prof. John Mattick and PhD student Mr Stefan Stanley, we developed an algorithm (GONOME) and statistical tools for analysing correlations between patterns in genomic DNA and the Genome Ontology (GO) annotation of the corresponding genes.
- (i) With visiting researcher Mr Rolv Seehuus, we investigated the application of data on positional distribution of DNA motifs, such as transcription factor binding sites, to the discovery of motifs in genomic sequence.

2.3 New statistical approaches in bioinformatics

Investigators:

Prof. Geoff McLachlan
(The University of Queensland)

Dr Markus Hegland
(The Australian National University)

In collaboration with:

Dr Christophe Ambroise
(Université de Technologie de Compiègne)

Prof. Kim-Anh Do
(M.D. Anderson Cancer Center,
University of Texas)

Dr Richard Kerr
(MMI Genomics)

Dr Thriyambakam Krishnan
(Systat Bangalore, India)

Dr Angus Ng
(The University of Queensland)

Researchers:

Research Associates:

Dr Richard Bean
(The University of Queensland)

Dr Liat Jones
(The University of Queensland)

PhD students:

Ms Katrina Monico
(The University of Queensland)

Mr Justin Xi Zhu
(The University of Queensland)

Aims:

- (a) Develop statistical and computational techniques to analyse cellular transcription.
- (b) Provide statistical tools for analysing high-throughput data, focusing on microarray gene-expression data, but also applicable to proteomics data.

Achievements (2006):

- (a) We developed and published in *Bioinformatics* a paper on the problem of differential expression that is applicable to data obtained under various experimental designs, including longitudinal and cross-sectional designs. Work is continuing on a user-friendly software program in R that implements this approach.
- (b) We completed the writing up of two book chapters on the clustering of microarray gene-expression data to be published respectively by Wiley and by Humana Press.
- (c) We continued with work on the use of gene-expression data in the construction of genetic regulatory networks (GRNs).
- (d) We completed the writing of the second edition of a Wiley monograph on the EM algorithm and its extensions. This monograph is used extensively in programs for the analysis of data in bioinformatics.
- (e) We presented an invited workshop on "Statistical methods for the screening and classification of microarray gene expression data" organized by the University of Waikato in Hamilton, New Zealand.
- (f) Together with colleagues from the MD Anderson Cancer Center (Houston, USA) we presented a workshop "Methods and computational tools for the screening and classification of microarray gene expression data" at the Joint Statistical Meetings in Seattle.



2.4 Mathematical modelling of gene regulatory networks

Investigators:

Prof. Kevin Burrage
(The University of Queensland)

Dr Markus Hegland
(The Australian National University)

In collaboration with:

Dr Pamela Burrage
(The University of Queensland)

Dr Margherita Carletti
(Università di Urbino, Italy)

Dr Grant Lythe
(University of Leeds, UK)

Researchers:

Principal Research Fellow:

Dr Jim Hanan
(The University of Queensland, jointly with
ACMC, ACCS and ACILR)

Research Associates:

Dr Nicholas Hamilton
(The University of Queensland)

Dr André Leier
(Advanced Computational Modelling Centre,
The University of Queensland)

Dr Lucia Santoso
(The Australian National University)

Dr Jiangning Song
(Advanced Computational Modelling Centre,
The University of Queensland)

PhD students:

Mr Shev MacNamara
(The University of Queensland)

Mr Dan Nicolau Jr
(The University of Queensland)

Aims:

- (a) Develop quantitative mathematical models to analyse biological systems and networks, in particular genetic regulatory networks, using stochastic differential equations and other formalisms, and implement them in computer code.
- (b) Test these models using computational simulation with both simulated and empirical data, and collaborate in their experimental validation where appropriate.
- (c) Collaborate and build synergies with other ARC research centres in mathematical modelling of biological systems at the cell and molecular levels.

Achievements (2006):

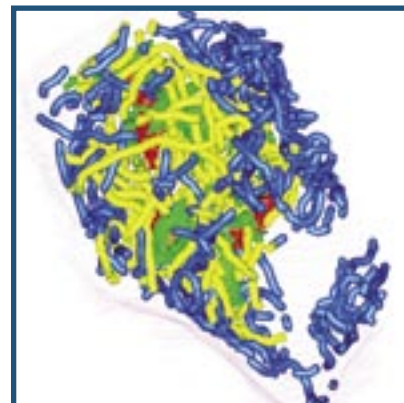
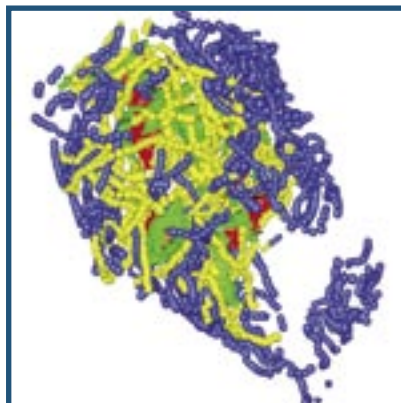
- (a) During 2006 we extended our models of genetic regulatory systems to incorporate external noise, and transcriptional and translational time delays. In addition we developed a new theory for the chemical master equation to accommodate delays in the response of cellular processes. This work appeared in *PLoS Computational Biology*.
- (b) We developed new techniques for simulating delay and noise in transcription and translation. This work appeared in a special Issue of *Journal of Computational and Applied Mathematics*.
- (c) We modelled delays in transcription and translation in genetic regulation associated with the *hes1* gene in mouse. This work was published in *PLoS Computational Biology*, and is the first publication of stochastic delay models that can explain experimental observations on the *hes1* molecular clock.
- (d) We developed and implemented stochastic models for regulatory networks of the genetic toggle switch. This work, which explains experimental observation of bimodal population distribution in proliferating bacteria, appeared in the *Proceedings of the National Academy of Sciences USA*.

- (e) We developed new methods for the efficient solution of the chemical master equation based on fast Krylov subspace methods for computing the exponential of a matrix times a vector. This work appeared in a special issue celebrating the 150th anniversary of Markov.
- (f) We developed new approaches based on the Simulated Maximum Likelihood method for estimating kinetic rates in gene expression. This work was published in *Bioinformatics*. The research represents an important step in deriving stochastic genetic regulatory models from experimental data.
- (g) We developed new frameworks, algorithms and software that position us to evolve genetic regulatory models from first principles. This work was presented at EuroGP 2006.
- (h) We developed new techniques for the prediction of *cis/trans* isomerisation in proteins using Psi-BLAST profiles and secondary structure information, and for predicting residue-wise contact orders in proteins by support vector regression. We also explored synonymous codon usage preferences of disulfide-bonded and non-disulfide bonded cysteines in the *Escherichia coli* genome. This work appeared in the *Journal of Theoretical Biology* and in *BMC Bioinformatics*.
- (i) We initiated new collaborations with researchers in Spain, the UK, and China. These relationships helped us forge new funding opportunities in the European 6th and 7th Framework rounds.

(j) We developed new numerical techniques using sparse grids and Dr Roger Sidje's Expokit to solve the master equations for gene regulatory networks for transcription and translation. Using these techniques, we have computed evolving probability distributions for these networks with up to 10 proteins. These distributions provide, for the first time, comprehensive information about the effects of randomness and low copy number on regulatory networks of this size. This work was done in collaboration with members of Centre for Bioinformation Science at ANU, and appeared in a special issue of the *Journal of Computational and Applied Mathematics*.

(k) We established error bounds for the approximation errors of the new sparse grid technology to solve the master equations started in (j). This work has been presented at the International Conference on High-Performance Scientific Computing Processes 10-24 March 2006, Hanoi, Vietnam and has been accepted to appear in the proceedings.

(l) We developed and implemented an adaptive approach for solving the master equations of gene regulatory networks with up to 100 proteins, using a totally new algorithm that combines ideas from graphical models and sparse grid approximation. This work was done at ANU with M.Sc. student Mr Leo Huber, who was awarded the MSc degree by his institution (ETH Zürich) at the end of 2006. During 2006 Mr Huber presented results in three seminars in Australia and Switzerland, and CI Markus Hegland will present this work at the Second Workshop on High-dimensional Approximation (HDA07) in February 2007. A manuscript has been written and will be submitted in early 2007.



2.5 (formerly 2.6) Complex systems modelling of regulatory and protein-protein interaction networks

Investigator:

Prof. Janet Wiles
(The University of Queensland)

In collaboration with:

Prof. John Mattick
(The University of Queensland)

Researchers:

Postdoctoral research fellow:

Dr James Watson
(The University of Queensland)

PhD students:

Mr Nic Geard
(The University of Queensland)

Mr James Watson
(The University of Queensland)

Mr Kai Willadsen
(The University of Queensland)

Research Assistant:

Mr Daniel Bradley
(The University of Queensland)

Honours student:

Mr Mark Wakabayashi
(The University of Queensland)

Aims:

- (a) Develop the Artificial Genome Algorithm as an abstract representation for the modelling of genetic regulatory networks (GRNs) in mammalian cells.
- (b) Apply insights gained from computational modelling to GRNs based on biological data, particularly those involving RNA control networks and cancer.
- (c) Quantify structural features of GRNs relevant to their behaviour and dynamics in living systems, by application of techniques of graph theory and analytical mathematics.

Achievements (2006):

- (a) The feature set of the Network Analyser has been finalised. The source and binaries are available from a SourceForge project web site (complexity.sf.net) that we created to enable ongoing development and collaboration. This site will be linked to from the ACB tools web site once the site design has been finalised. The web site will provide access to documentation covering the use of the tools and of the associated development libraries.
- (b) A software suite tying together computational models of genetic sequence, regulatory network, plant development and evolution has been completed and published (James Watson's PhD thesis, 2006). Development of a reusable library of in-house modeling software was initiated in collaboration with the ACCS (summer studentship) to make this work available to the wider modelling community.
- (c) An analysis of Boolean models of the *Drosophila melanogaster* segment polarity network and the *Saccharomyces cerevisiae* cell-cycle network was completed and published (Kai Willadsen's PhD thesis, 2006). This study demonstrated that, while both systems are robust, the structure of the state space plays a variable role in system robustness, with the yeast network displaying a much less structured state space than the *Drosophila* network.

- (d) We continue to investigate how genetic regulatory networks control developmental processes, by testing the ability of artificially evolved networks to reproduce target information patterns such as the differentiation decisions for the cell lineage of *Caenorhabditis elegans* and other complex systems. A model of artificial ontogenies was completed and published (Nic Geard's PhD thesis, 2006), which incorporates a mapping between the dynamics of a gene network and a cell lineage representation of a phenotype. The thesis explored ontogenetic complexity and its relationship to the genetic regulatory network architecture of developmental control. This work won the CORE Distinguished Doctoral Dissertation Award for the best Australian computer science thesis in 2006.

2.6 (formerly 2.7) Ortholog mapping of prokaryotic and eukaryotic genomes

Investigator:

Prof. Mark Ragan
(The University of Queensland)

In collaboration with:

Dr Robert Charlebois
(Genome Atlantic, Canada)

Prof. Mike Fellows
(University of Newcastle)

Dr Jonathan Keith
(Queensland University of Technology)

Dr Fran Rosamond
(University of Newcastle)

Researchers:

Research Associates:

Dr Robert Beiko
(The University of Queensland)

Dr Aaron Darling
(The University of Queensland)

Dr Karin Kassahn
(The University of Queensland)

Dr Simon Wong
(The University of Queensland)

PhD students:

Mr Cheong Xin Chan
(The University of Queensland)

Mr Michael Höhl
(The University of Queensland)

Mr Chang Jin Shin
(The University of Queensland)

Aims:

Homology is the foundation of comparative genomics, and orthology is its most solid cornerstone. Identifying orthologs (regions that are precise evolutionary counterparts in different genomes) is fundamental to understanding how chromosomes, gene and protein families, molecular networks and thus cellular functions evolve. Cryptic or degenerate control signals can be uncovered most effectively (for example using pattern discovery, hidden Markov models and other advanced statistical tools) by comparing orthologous regions. In prokaryotes, where some genomic regions have been transferred across, not along, genealogical lineages, ortholog mapping helps us identify *prima facie* instances of lateral genetic transfer (LGT).

- Consolidate the development of an automated high-throughput bioinformatic pipeline that identifies protein family clusters over multiple genomes, generates optimal multiple sequence alignments, infers statistically sound phylogenetic trees, and rigorously compares subtrees to identify instances and patterns of topological congruence among trees. These instances of incongruence are *prima facie* cases of lateral genetic transfer (LGT).
- Continue to apply this bioinformatic pipeline to identify orthologous and paralogous gene families, ortholog-characteristic patterns, and *prima facie* instances of LGT among prokaryotic genomes.
- Characterise these instances of LGT statistically, and using advanced database methods correlate them with genomic, physiological, metabolic, ecological, phyletic and other features to develop and test specified hypotheses about the frequency, mechanisms, patterns and biological significance of LGT among prokaryotes.
- Continue to develop and implement additional capabilities into this pipeline, notably algorithms and tools that allow the automated recognition and analysis of arbitrarily or self-defined sub- and super-genic regions.
- Develop a Grid-aware portal to manage this pipeline, including genome and protein sequences, clusters, alignments, trees, comparisons, analytical results and metadata.

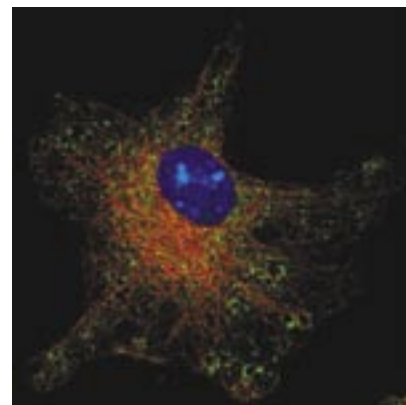
- Define needs and develop technical specifications for similar automation of comparative genomic analyses over additional data domains, including partial genomic and EST sequences, environmental genomic data and complex eukaryotic genomes.

Achievements (2006):

- We completed development of the core high-throughput bioinformatic pipeline for LGT. All main algorithmic components have now been published, and an architectural overview is in press. In September Dr Rob Beiko accepted a junior Canada Research Chair at Dalhousie University.
- Using our rigorous tree-based results over 144 prokaryotic genomes, we demonstrated that different statistical approaches preferentially detect LGT events of different relative ages (*Trends in Microbiology*). We applied our pipeline to two specific prokaryotic genome datasets, for genus *Staphylococcus* and for cyanobacteria; manuscripts are being prepared.
- PhD student Mr Cheong Xin Chan completed development of a two-stage module that identifies regions of genetic recombination and estimates the position of recombination breakpoints under different statistical models. We benchmarked this extensively (*BMC Bioinformatics*), implemented key portions on the APAC national supercomputer facility, and applied it to the *Staphylococcus* genome dataset. This now makes it possible for us to examine the relationship between genetic recombination and LGT, and to test hypotheses about the frequency, mechanisms, patterns and biological significance of both genetic recombination and LGT. With the recruitment of Dr Aaron Darling (NSF Fellow) into the project, we implemented his multiple whole-genome alignment package MAUVE and undertook a preliminary assessment of its utility in our pipeline.
- In view of the establishment of the Queensland Facility for Advanced Bioinformatics (QFAB) and capability within QFAB for a generic approach to bioinformatic workflow portals, we postponed development our own portal for managing the bioinformatic pipeline described above.



- (e) The sub-project on alignment-free comparative methods was concluded with acceptance of two very substantial manuscripts (*Evolutionary Bioinformatics Online* 2006, *Systematic Biology* 2007 to appear) and a PhD completion for Mr Michael Höhl. The former paper was co-authored with ACB PI Dr Isidore Rigoutsos.
- (f) During 2006 the main focus of Project 2.7 shifted from LGT and recombination in prokaryotic genomes, to genomic and cellular determinants of conservation among large eukaryotic genomes. Dr Simon Wong continued development of algorithms and software for the identification at high resolution of orthologous and paralogous regions among the human, mouse and other large complex eukaryotic genomes. Collaboration with ACB CI Prof. Mike Fellows led to a novel algorithmic analysis and the submission of three manuscripts (*Texts in Algorithmics* 2006, plus two under review).
- (g) Dr Karin Kassahn was recruited to the project, and will apply our bioinformatic pipeline to identify orthologous and paralogous regions among fish genomes, with the goal of identifying fast- and slow-evolving genomic regions to compare their evolution pattern and mode. We took part in a successful ARC LIEF application, led by ACB Advisory Board member Prof. John Mattick, for infrastructure in support of an Australian mirror of the UCSC Genome Browser.
- (h) In collaboration with ARC Fellow Dr Jonathan Keith, we continued foundational work on a general statistical description of evolutionary processes underlying the comparison of large eukaryotic genomes.
- (i) PhD student Mr Chang Jin Shin substantially completed the construction of an integrated database linking positional, protein-protein interaction, and subcellular localisation data in human and mouse. We related these data among themselves and with underlying evidence based on analytical technology, and initiated two manuscripts. To the extent possible, we use the same description vocabulary (based on Gene Ontology) as the LOCATE database (Project 1.2) and the Visible Cell™ environment (Program 3). We developed a principled approach to describing adjacency in subcellular space, and initiated a manuscript.



Program 3 – The Visible Cell – Data integration and visualisation

3.1 Advanced data management and integration

3.1.1 Data environment for the Visible Cell™ project

Investigators:

Prof. Mark Ragan
(The University of Queensland)

Prof. Xiaofang Zhou
(The University of Queensland)

In collaboration with:

Prof. Kaye Basford
(Australian Centre for Plant Functional Genomics)

Ms Kimberly Begley
(Griffith University, and APAC Grid Bioinformatics Program)

Dr Lindsay Hood
(The University of Queensland)

Dr Ralf Muhlberger
(The University of Queensland)

Prof. Bernard Pailthorpe
(The University of Queensland)

Researchers:

Research Associate:

Dr Lynn Fink
(The University of Queensland)

PhD students:

Mr Alex Garcia
(The University of Queensland)

Ms Helen Zi Huang
(The University of Queensland)

Research staff:

Mr Oliver Cairncross
(ARC Centre in Bioinformatics)

Ms Mhairi Marshall
(ARC Centre in Bioinformatics)

Mr Ken Steube
(ARC Centre in Bioinformatics, and APAC Grid Bioinformatics Program, 1/06-5/06)

Aims:

- (a) Develop techniques and implement databases that deal powerfully with modern genome-phenome data, including molecular sequences, structures, pathways, molecular interaction and regulatory networks, results of computational modelling and simulation, images, data from array-based phenomic research, automatic annotations and metadata.
- (b) Identify, implement and integrate technologies that make it possible to deliver these data into a common research and visualisation environment (the Visible Cell™) in an efficient, time-critical manner. These include generic and domain-specific technologies for data management and integration, mass storage, indexing and retrieval, as well as XML and specialised mark-up languages.
- (c) Develop fast database indexing methods to support efficient biological query processing with very large amount of complex biological data, in particular focusing on protein surface docking problems.
- (d) Grid-enable the Centre's research databases and selected data.
- (c) We implemented formatting, auto-tiling and resampling capabilities within the Visible Cell™ data store, such that terabyte raw data from electron microscopes can be efficiently prepared *in situ* for downstream processing within the Visible Cell™ visualiser. We implemented code for input of image data in .mrc format (for EM and tomography data) and export in .png format. We investigated the Zeiss (light microscope) format but have not yet implemented code to read it. The processed data form the working matrix of our Visible Cell™.
- (d) A mirror of the multi-genome database ENSEMBL was successfully installed at Griffith University, with both production (open) and development (closed) versions in place. The production mirror currently serves ENSEMBL release 41 (<https://sunfire03.rcs.griffith.edu.au:8443/index.html>). Scripts were implanted for automatic data download, back-end data builds and the front-end web interface. It remains to get SRB and the ENSEMBL API talking to each other.
- (e) With the funding of the Queensland Facility for Advanced Bioinformatics (QFAB), we began migrating much of the Centre's infrastructure development (including SRB, CCDB, the ENSEMBL and UCSC Genome Browser mirrors, and PDB) to QFAB (page 40). QFAB intends to licence-in commercial software that will automate many routine tasks in the management of large bioinformatic databases.
- (f) We developed a multi-resolution local occurrence model (LCM) which considers the neighbourhood information of 3D structures to support efficient sub-structure similarity search. This method has proven to be much more efficient than existing approaches for 3D structure similarity search, including geometric hashing and DALI. We have applied this model to protein structure analysis, and generalised it to support other types of 3D molecular structural similarity search. We also developed a summarisation-based method, BCS, which applies principal component analysis to compress three- and high-dimensional point data into a highly compact coordinate system for structure similarity search.

Achievements (2006):

- (a) The Grid middleware application Storage Resource Broker (SRB) version 3.4.2 was installed within ACB (Institute for Molecular Bioscience) and at Griffith University, and is available to users at both sites. SRB enables access control and sharing of very large heterogeneous datasets across a Grid.
- (b) Cell Centered Database (CCDB), a Grid middleware application specialised for management of 2D and 3D biological images and associated metadata, was installed within ACB but the release version remains unstable. We built close links with the development team at the National Center for Microscopy and Imaging Research (NCMIR) headed by Prof. Mark Ellisman at UCSD, and are working closely with them to resolve stability issues and integrate functionality that will be required by the Visible Cell™ project. It is our intention to make CCDB accessible *via* SRB. ACB/IMB is only one of two alpha test sites worldwide.



- (g) PhD student Mr Alex Garcia submitted his thesis on ontology development in molecular bioscience. Results from his internship at the European Bioinformatics Institute (see our 2005 Annual Report, page 53) were published in *BMC Bioinformatics* and in a paper at the *Protégé-9* conference. Alex accepted a data integration position with Centro Internacional de Agricultura Tropical in his native country Columbia, and has been posted by CIAT to Stanford University.
- (h) ACB was represented in a consortium (ANU, Monash University, Queensland Cyber Infrastructure Foundation, UQ) that was awarded ARC LIEF funding to implement the access layer for an integrated scientific data storage capacity for Australia.
- (i) Expertise developed in this Project and in Project 3.2.2 was instrumental in development of a research collaboration with Pfizer Research Technology Center (Cambridge, MA USA). The research contract was executed on 21 December 2006, with work to begin in early 2007.
- (j) We continued to track Australian e-research funding opportunities, with the intention of securing additional resources to develop the Visible Cell™ beyond prototype.



3.1.2 Integrating database technologies and visual analysis for genome data

Investigators:

A/Prof. Yi-Ping Phoebe Chen
(Deakin University)

Prof. Xiaofang Zhou
(The University of Queensland)

In collaboration with:

Dr Timothy Bailey
(The University of Queensland)

Prof. Kevin Burrage
(The University of Queensland)

Dr Sean Grimmond
(The University of Queensland)

A/Prof. Pablo Moscato
(The University of Newcastle)

Prof. Mark Ragan
(The University of Queensland)

Researchers:

Research Associate:

Dr Jiyuan An
(Deakin University)

PhD students:

Ms Yan Chen
(Deakin University)

Mr Alex Garcia
(The University of Queensland)

Mr Keith Knapp
(Deakin University)

Aims:

- (a) Develop and refine data-mining techniques, particularly the selection of effective dimensions, for discovery of common emerging patterns among gene expression data.
- (b) Apply these methods to classify gene expression patterns from normal and diseased tissues.
- (c) Apply these methods to data on gene expression in response to trace metals, to discover regulation patterns and gene-regulatory motifs.
- (d) Interrelate disparate data types (sequences, structures, expression data) by classifying and clustering genomic data using data mining and machine learning technologies.
- (e) Link interactive methods with data-mining techniques to contribute to the Visible Cell™ environment.

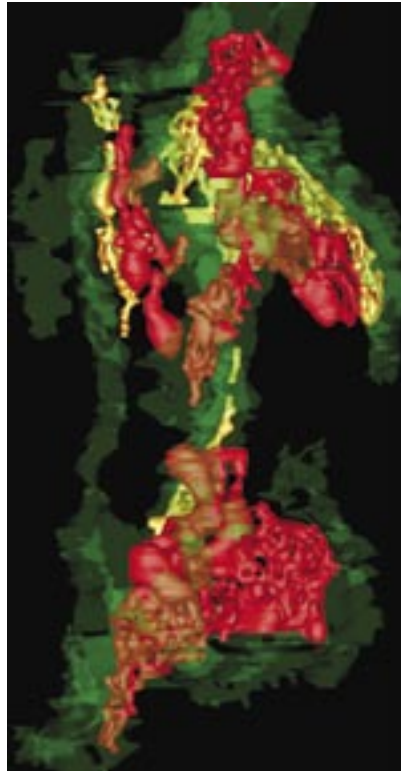
Achievements (2006):

- (a) Pattern recognition is a classical topic in information science. Traditionally, patterns are defined as appearing in one class but not in any other. However, the patterns in genomic data more typically appear in some examples of one class but not in any other class. For example, in a prostate cancer gene expression dataset, we found three patterns that appear in 72.7% of patients but not in the controls, and 129 patterns that appear in 67.5% of patients but not in controls. This kind of pattern is called an emerging pattern. We published these results in the proceedings of the 18th International Conference on Pattern Recognition (ICPR2006).
- (b) We developed a new classification method based on “divide and conquer” in which junk dimensions can be efficiently removed, with the effective dimensions identified at the early stages. We also developed and implemented a depth-first approach to pattern discovery, similarity matching and visualisation for genomic sequence data that likewise speeds up pattern discovery by removing junk dimensions at early stages of the analysis, and submitted a paper on it at the Fourth International Conference on Active Media Technology (AMT 2006).

- (c) We continued work on dimensionality reduction in protein structure matching using motif signatures, and published a paper in the 17th Australasian Database Conference (ADC 2006). Work on protein structure matching using patch signatures was published in the 17th International Conference and Workshop on Database and Expert Systems Applications (DEXA 2006).
- (d) Short patterns are often more interpretable biologically (*e.g.* as molecular binding sites). We developed and implemented a new algorithm to find short patterns from multi-class data, and used 10-class dataset to evaluate its performance, with encouraging results. This work was published in the IEEE Web Intelligence conference (WI2006).
- (e) We continued to examine the effect of trace metals such as copper on patterns of gene expression, and presented a conference paper on our work at AMT 2006.
- (f) To assist in the analysis of very large molecular-sequence data, we proposed an interactive 3D visualisation model in which end-users select the dimensions in which to display the data distribution. A paper on this work was accepted in *Journal of Visual Languages and Computing*. We began to apply this approach to gene expression datasets.
- (g) In collaboration with A/Prof. Pablo Moscato, we have implemented a method to discriminate diseased from normal tissues. Working with 2100 gene expression data points from 24-voxel images of coronal hemisections of both normal and diseased human brain, we applied machine learning methods to locate genes or groups of genes correlated with the expression of Alzheimer’s disease.



- (h) We proposed a framework that can assist in the identification of potential correlation among genes. We applied item constraints in the closed interpretation to itemset generation so that a threshold is specified in terms of the amount of results, rather than a fixed threshold value for all itemsets of all sizes. We analysed the rules derived in this way from data on the regulation of AMP-activated protein kinase, finding that most of the extracted association rules have biological meaning, including some rules that were previously unknown. This work was published in *BMC Bioinformatics*.
- (i) Outcomes from collaboration during 2005 with colleagues at the European Bioinformatics Institute appeared in *BMC Bioinformatics*. In this paper we report the comparison, development and extension of methodologies for eliciting knowledge from geographically disperse members of an expert community, formally mapping concepts, and generating ontologies in bioscience. We specified user needs for software to support these processes, and contributed to the development of prototypes. The PhD student involved in this collaboration, Mr Alex Garcia, submitted his thesis for examination.



3.2 Visualisation

3.2.1 Concept and engineering of the Visible Cell™

Investigators:

Prof. Mark Ragan
(The University of Queensland)

Dr Brad Marsh
(The University of Queensland)

Project Manager:

Dr Lindsay Hood
(The University of Queensland, 1/06-6/06)

Mr Oliver Cairncross
(ARC Centre in Bioinformatics, from 7/06)

In collaboration with:

Dr Ben Hankamer
(The University of Queensland)

Prof. Jane Hunter
(The University of Queensland)

Prof. Xiaofang Zhou
(The University of Queensland)

Researchers:

Principal Research Fellow:

Dr Jim Hanan
(The University of Queensland, jointly with
ACMC, ACCS and ACILR)

Research Associate:

Dr Nicholas Hamilton
(The University of Queensland)

ACB staff:

Mr Matthew Bryant
(ARC Centre in Bioinformatics)

Mr Igor Kromin
(ARC Centre in Bioinformatics)

Mr Tim McComb
(ARC Centre in Bioinformatics)

Ms Mhairi Marshall
(ARC Centre in Bioinformatics)

Mr Ken Steube
(ARC Centre in Bioinformatics, and APAC
Grid Bioinformatics Program, 1/06-5/06)

Mr David Wood
(ARC Centre in Bioinformatics)

Honours student:

Mr Robert McLeay
(ARC Centre in Bioinformatics)

Aims:

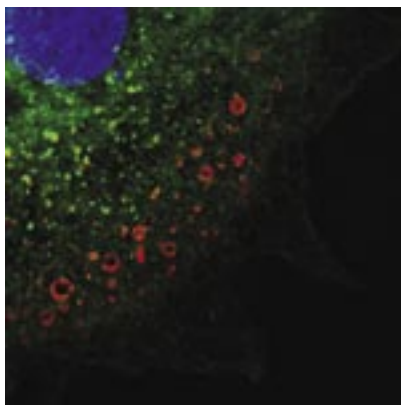
- (a) Refine our conceptualisation of a three-dimensional visualisation environment for exploring the biology of the mammalian cell that seamlessly integrates empirically determined molecular structures and interactions, quantitative simulations based on mathematical and complex systems models, and high-resolution cell tomograms and electron micrographs. This environment, Visible Cell™, will support user interaction with the 3-D cellular models, facilitating hypothesis-generation and hypothesis-testing in genome-phenome biology including gene regulation, cellular growth and development, metabolism, physiology, transport, cell-cell interaction, disease, and cell death.
- (b) Based on technical specifications developed earlier, undertake modular development of a prototype Visible Cell™.
- (c) Explore in detail whether the Cell Illustrator environment can provide the software basis for a pathway modelling and visualisation module within our Visible Cell™.
- (d) Deepen collaboration with the Data Integration project (3.1, above) and other activity in ACB with the goal of basing the Visible Cell™ environment over federated databases that will support hypothesis-generation and hypothesis-testing by the efficient delivery of semantically rich information, including user-specified metadata.
- (e) Collaborate with experimental phenomics researchers in the use of this environment in the generation and testing of biological hypotheses in mammalian cell biology.

Achievements (2006):

- (a) During 2006 we focused Centre resources on the Visible Cell™ project to ensure a successful transition from design specifications and initial prototypes, and development of momentum leading to modules of increasing capability, functionality and flexibility. With the departure of Dr Lindsay Hood to APAC, Mr Oliver Cairncross was named Project Manager. Mr Tim McComb was recruited as application developer after Mr Igor Kromin went to Oracle. Mr David Wood (databases) was recruited from Australian Genome Research Foundation. Mr Tim Sullivan and Dr Muhammad Shoaib Sehgal will join the project in early 2007 from QUT and Monash University respectively.
- (b) Through IMBcom, we filed for registration of Visible Cell™ jointly with UQ as an international trademark in Australia, New Zealand, the USA, Europe and the United Kingdom.
- (c) We launched a public website (<http://www.visiblecell.com>). This supplements a large internal project wiki.
- (d) We improved the data integration component of Visible Cell™ to allow terabyte image datasets to be uploaded and processed in a single pass. This has greatly increased the speed at which image data are loaded into the system. With respect to visualising surface reconstructions: the data transfer speed to the visualiser has been increased; the user interface for interacting with the data has been made more flexible; and the user has been given more options for presentation and formatting.
- (e) We re-engineered the internal command structure of .CELL (dot-cell) to make it more easily extensible and to improve its ability to host plug-ins. These changes made it easier for us to implement the achievements reported in (f) and (g) below.



- (f) We expanded the .CELL platform to allow other applications to interact with the Visible Cell™ environment via simple commands in a client-server framework. This will allow biologists to interact with the data using techniques with which they are familiar (e.g. a Perl script). We also took the first steps to enable external applications, including Cell Illustrator™, to execute models using Visible Cell™ data with the results displayed in the Visible Cell™ visualiser. An unexpected side effect of this development was that we can now link multiple visualisers together, opening interesting possibilities for facilitating collaboration among remote researches, and for using the system for teaching cell biology or systems modelling.
- (g) Leveraging the Centre's computational cellular biology research (Project 1.2) and resources within QFAB (page 40) yielded significant progress toward integrating federated data on protein structure, function and interactions with the Visible Cell™ structural data. We developed a mechanism that allows the user to assign controlled terms from Gene Ontology (as extended by the LOCATE database) to subcellular components (volumes or surfaces) in the visualised cell data. We are able to use these terms to retrieve further information from federated databases using Sequence Retrieval System (SRS).
- (h) We commenced work on visualising the results of cellular pathway queries. The database schema has been expanded, and data-loading scripts have been written and incorporated. The query engine and user interfaces are in the prototype phase.



3.2.2 (incorporating the former 2.5) Spatial modelling of cellular systems

Investigators:

Prof. Kevin Burrage
(The University of Queensland)

Dr Rohan Teasdale
(The University of Queensland)

Dr Timothy Bailey
(The University of Queensland)

Dr Markus Hegland
(The Australian National University)

In collaboration with:

Prof. John Hancock
(The University of Queensland)

A/Prof. Rob Parton
(The University of Queensland)

Prof. Jenny Stow
(The University of Queensland)

Researchers:

Principal Research Fellows:

Dr Jim Hanan
(The University of Queensland, jointly with
ACMC, ACCS & CILR)

Dr André Leier
(Advanced Computational Modelling Centre,
The University of Queensland)

Dr Jiangning Song
(Advanced Computational Modelling Centre,
The University of Queensland)

Research Associate:

Dr Nicholas Hamilton
(The University of Queensland)

PhD students:

Mr Shev MacNamara
(The University of Queensland)

Mr Dan Nicolau Jr
(The University of Queensland)

Aims:

- (a) Develop active collaborations among mathematicians, modellers and experimental biologists enabling the realistic quantitative description of selected structures, spatial relationships and dynamic processes in mammalian cells.
- (b) Develop methodology and local expertise in the modelling of biological processes, with particular focus on the application of stochastic simulators to lipid rafts and anomalous diffusion.
- (c) Develop or modify generalised codes that can be rapidly configured for specific simulation requirements, use these tools in the development of higher-level formalisms, and incorporate them into simulators as requirements are determined.
- (d) Work toward the collaborative experimental validation and refinement of these models.

Achievements (2006):

- (a) We interrogated a stochastic spatial model of the plasma membrane to determine how micro-domains affect protein dynamics. Taking lipid rafts as a representative micro-domain, we showed that reduced protein mobility in rafts segregates dynamically partitioning proteins, but the equilibrium concentration is largely independent of raft size and mobility. Rafts weakly impede small-scale protein diffusion, but more strongly impede long-range protein mobility. The long-range mobility of raft-partitioning and raft-excluded proteins is reduced to similar extents. Dynamic partitioning into rafts increases specific inter-protein collision rates, but to maximise this critical, biologically relevant raft function, rafts must be small (diameter 6-14 nm) and mobile. Intermolecular collisions can also be favoured by the selective capture and exclusion of proteins by rafts, although this mechanism is generally less efficient than simple dynamic partitioning. Thus micro-domains can readily operate as protein concentrators or isolators, but with significant constraints on

size and mobility if micro-domains are also required to function as reaction chambers that facilitate protein-protein interactions. Our results may have significant implications for signalling cascades that are scaffolded or assembled in plasma membrane micro-domains. This work appeared in *Molecular and Cellular Biology* and is one of the most sophisticated computational models of the cellular plasma membrane.

- (b) We developed a stochastic random walk model of protein molecule diffusion on a cell membrane, and investigated the fundamental causes of anomalous diffusion in two-dimensional biological media. Three different interactions were considered: collisions with fixed obstacles, "picket fence posts", and capture by (or exclusion from) lipid rafts. If motion is impeded by randomly placed, fixed obstacles, diffusion can be highly anomalous. In contrast, collision with picket fence posts has a negligible effect on the anomalous exponent at realistic parameters. The effects of lipid rafts are more complex. If proteins partition into lipid rafts, there is a small to moderate effect on the anomalous exponent, whereas if proteins are excluded from rafts there is a large effect on the anomalous exponent. In combination, these mechanisms can explain the level of anomaly in experimentally observed membrane diffusion, suggesting that anomalous diffusion is caused by multiple mechanisms whose effects are approximately additive. Finally, we showed that the long-range diffusion rate D_{macro} , estimated from Fluorescence Recovery After Photobleaching studies, can be much smaller than D_{micro} , the small-scale diffusion rate, and is highly sensitive to obstacle densities and other impeding structures. This work appeared in *Biophysics*.



- (c) We used Monte Carlo simulations to estimate an anomalous diffusion parameter that encapsulates the crowdedness of the spatial environment. We then used this parameter to replace the rate constants of bimolecular reactions by a time-dependent power law to produce an SSA valid in cases where there is anomalous diffusion, or where the system is not well-mixed (ASSA). Simulations show that ASSA can successfully predict the temporal dynamics of chemical kinetics in a spatially constrained environment. This work appeared in a special issue of *Computers and Mathematics with Applications*.
- (d) In collaboration with Professor Jenny Stow and her group (IMB) we have observed apparent periodicities in the timing of tubulation events in the Golgi and are using computational modelling to examine the phenomenon in more depth. A paper on this work will appear in *Current Protocols in Cell Biology*. We continued development of simple geometric models that allow us to explain the dynamics and geometry of vesicles and tubules in endocytosis.
- (e) We initiated collaboration with A/Prof. Rob Parton (IMB) with the goal of modelling the mobility of lipid droplets in the cell. This will yield information more broadly on spatial and temporal dynamics in the cytoplasm. We are automating the analysis using video imaging under a range of experimental conditions.
- (f) We developed stochastic mathematical techniques to interpret observations and experimental results, allowing us to build simple stochastic genetic regulatory models from experimental data.
- (g) In another project directed to modelling genetic regulation, we used the Cell Illustrator™ software to elaborate a Petri net model of Toll-like receptor pathways in mammalian cells. We refined the model in a series of discussions with Prof. David Hume and his research group (ARC SRC for Functional and Applied Genomics). Robert McLeay made a more general presentation of the Cell Illustrator™ modelling system to the SRC. After trials, the SRC group decided to use a simpler approach.
- (h) Discussions were held with Dr Brad Marsh's group (IMB) to determine networks for further modelling using the Petri net / spatial modelling approach. A pilot project combining Petri nets with rule-based spatial models was completed by student intern Fritjof Heyde, with intracellular aspects of the mammalian beta cell insulin production and release cycle proposed as the target system.
- (i) We are coordinating with the Visible Cell™ project team to build an increasingly detailed work plan of how these simulators can be attached to the Visible Cell™ prototype, and have developed strategies for linking biochemical simulators to visualisations within the Visible Cell™.
- (j) We are also developing methodologies and software to provide more refined inference of protein co-localisation 4D.
- (k) Expertise developed in this Project and in Project 3.1.1 was instrumental in development of a research collaboration with Pfizer Research Technology Center (Cambridge, MA USA). The research contract was executed on 21 December 2006, with work to begin in early 2007.

3.3 Bioinformatics web services (completed, 2006)

Investigator:

Prof. Mark Ragan
(The University of Queensland)

In collaboration with:

Dr Kimberly Begley
(Griffith University, and APAC Grid Bioinformatics Program)

Dr Lindsay Hood
(The University of Queensland 1/06-6/06, and Australian Partnership in Advanced Computing 7/06-12/06)

Researchers:

Research Associate:

Dr Lynn Fink
(The University of Queensland)

ACB staff:

Ms Mhairi Marshall
(ACB and APAC Grid Bioinformatics Program 6/06-12/06)

Mr Ken Steube
(ACB and APAC Grid Bioinformatics Program 1/06-5/06)

Aims:

- (a) Improve user access to tools and facilities at the bio-IT interface.
- (b) Implement advanced information technologies that will enable the automated discovery of bioinformatic web tools and data sources.
- (c) Participate in and collaborate with national initiatives in Grid bioinformatics, including the Australian Partnership in Advanced Computing (APAC) Grid bioinformatics programme.

Achievements (2006):

- (a) Development of the ACB research website (<http://bioinformatics.org.au>) continued. At year's end, 23 bioinformatics packages, tools and datasets were publicly available (many of them uniquely) through our research website, and a further 27 external resources were linked. Several of our tools access high-performance computing resources within ACB or our partner institutions. Further packages and datasets, including a mirror of the IBM Pattern Discovery webtools, are available within ACB.
- (b) ACB secured continued co-investment from APAC for development of Grid technologies in bioinformatics. Ms Mhairi Marshall took responsibility for the project within ACB after Mr Ken Steube was recruited away.
- (c) The final day of the 2006 Winter School in Mathematical and Computational Biology was devoted to the Grid paradigm in bioinformatics and computational biology (see page 49).
- (d) Implementation of the Grid middleware applications Storage Resource Broker (SRB) and Cell-Centered Data Base (CCDB), and the ENSEMBL multi-genome database, are reported under Project 3.1.1.
- (e) We were part of a successful ARC LIEF application, led by ACB Board member Prof. John Mattick, to secure funding to implement an Australian mirror of the UCSC Genome Browser (<http://genome.ucsc.edu/>). Management of this resource will be transferred to the new Queensland Facility for Advanced Bioinformatics (see QFAB page 40).
- (f) With the startup of the QFAB, most development of Web services, Grid technologies and associated tools within ACB (Projects 3.1.1 and 3.3) is being transferred to QFAB (see page 40). Project 3.3 has thus been completed. ACB remains deeply committed to research and leadership in advanced information technologies for bioinformatics, and in their availability to and use by Australian researchers.





Queensland Facility for Advanced Bioinformatics (QFAB)

Genomics and its relatives – transcriptomics, proteomics, metabolomics and the others – offer unprecedented power to modern biotechnology, but harnessing this power is not easy. Datasets are large and semantically rich; novel technologies yield unfamiliar data types; user expectations are diverse, dynamic and often ill-defined. Data capture, data lifecycle management, security, high-performance computing and software maintenance are resource- and expertise-intensive. Even large organisations find it difficult to recruit the right mix of skills and maintain the breadth of infrastructure needed; smaller companies, or institutions newly adding omics to their technology mix, may find the challenges insurmountable.

Let us unpack this a little more. A small or medium-sized company is unlikely to find any single university or research centre with the full range of expertise and infrastructure it needs for a given project. Across institutions, infrastructure and skills may be duplicated, incompatible, or simply too difficult to coordinate. Governments may be reluctant to support individual projects if they perceive that synergies or economies of scale are being missed.

The bioinformatics R&D community in southeast Queensland is the largest and most-capable in Australia. We're also a sociable mob, attending each others' seminars, co-supervising students, applying for grants together, and meeting at State Government biotechnology and ICT functions. We know many of each others' strengths – workflow engines at QUT, database mirrors at Griffith, EST annotation at DPI&F, data integration at eHRC, HPC and petabyte storage at APAC and QCIF. As the other side of this same coin, we

appreciate that no single institution can be a comprehensive one-stop shop for all the informatic, statistical and computational needs of our domestic biotechnology industries, or partner effectively to link omic bioinformatics across e-health, e-research and advanced ICT. This message is reinforced in discussions with our colleagues in the private sector.

NIRAP (National and International Research Alliances Program) is a competitive, merit-based program within the Queensland Government's Smart State strategy that offers matching funds in support of strategically directed projects. In May 2005 Dr Anthony Maeder (e-Health Research Centre, a joint venture of the Queensland Government and CSIRO) came to me with the idea of a consortium in advanced bioinformatics infrastructure and delivery; our first detailed discussion took place at that year's meeting of the ARC Research Network in Enterprise Information Infrastructure. With so much networking and good will already in place, and high-level support from all institutions, we proposed a Queensland Facility for Advanced Bioinformatics (QFAB) in the December 2005 NIRAP round. On 11 April 2006, speaking at BIO 2006 in Chicago, Queensland Deputy Premier the Hon. Anna Bligh MP announced funding for QFAB.

The initial partners in QFAB are The University of Queensland (including ACB and the SRC for Functional and Applied Genomics), Griffith University, Queensland University of Technology, e-Health Research Centre, Queensland Dept of Primary Industry and Fisheries, Australian Partnership in Advanced Computing, and Queensland Cyber Infrastructure Foundation. The initial start-up phase was led by Interim CEO Mr Jeremy Barker (formerly COO of Australian Genome Research Foundation) and Interim Technical Manager Dr David Hansen (eHRC), culminating in a successful public launch on 21 December 2006.

QFAB (<http://www.qfab.org.au>) provides its partners a secure environment with access to a bioinformatics databases, search tools, commercial and specialist software, computing, storage, data integration technologies and expert services. Eleven personnel have been seconded, most of them full-time, from partner institutions. For ACB, participation in QFAB helps us address the third point of our Mission – improving access to skills, tools and facilities – while empowering our researchers to focus on research outcomes.

QFAB is also developing a public component with access to a certain level of resources including data and web tools. Its critical mass, high-level skills base, superb data and computational infrastructure, active partner involvement, and State support make QFAB an attractive participant in existing or proposed research consortia, commercial projects, and national collaborative research infrastructure in bioinformatics.

Despite the "Q" word, QFAB aspires to earn a place on the national and international stages. NIRAP presupposes a national and international focus, and we're well aware that opportunities in biotechnology, biomedicine and ICT are increasingly global. It's too early for AFAB, much less IFAB; but watch this space!

Mark Ragan

Key Performance Indicators

ACB's Key Performance Indicators (KPIs) encompass the following areas:

1. Research findings
2. Research training and professional education
3. International, national and regional links and networks
4. End-user links
5. Organisational support
6. National benefit

Highlights for 2006:

ACB hosted and sponsored the 2006 Winter School and Mathematical and Computational Biology. Our co-sponsors were the ARC Research Network in Enterprise Information Infrastructure (EII), Queensland Parallel Supercomputing Foundation (QPSF, since renamed Queensland Cyber Infrastructure Foundation), The MathWorks Australia Pty Ltd, and the Institute for Molecular Bioscience (IMB). Our Winter School is designed for postgraduate and advanced undergraduate students, postdoctoral researchers, and others working in mathematics, statistics, computer science, information technology, and biological, chemical or medical sciences or engineering. The 2006 Winter School, our third in this annual series, was again highly successful, attracting 148 registrants including several science teachers from Queensland secondary schools.

A student symposium, sponsored jointly by ACB and Bioinformatics Institute New Zealand in Auckland from 11-14 July, attracted approximately 90 participants. All students were required to present their work, either as a short oral paper or as a poster. ACB was represented by four researchers and eight postgraduate students. Some students also attended satellite workshops of the Australia-NZ Statistical Association Joint Meeting immediately before the symposium, with BINZ generously covering their registration fees.

Research from our Investigators continues to appear in top journals. In 2006, papers from ACB researchers appeared in *Cell*, *Genome Biology* (four), *Genome Research* (two), *Nature* journals (four), *PLoS Genetics* (five), and *Proceedings of the National Academy of Sciences USA* (two). Our work also appeared in top bioinformatics journals including *Bioinformatics* (four), *BMC Bioinformatics* (nine), and *PLoS Computational Biology*.

On our behalf and jointly with UQ, IMBcom filed for registration of Visible Cell™ as an international trademark in Australia and New Zealand (September) and in the USA, Europe and the United Kingdom (October). A public website (<http://www.visiblecell.com>) was launched in October.

Dr Rohan Teasdale's LOCATE database was featured in the 14 July 2006 issue of *Science*. LOCATE provides a unique multimedia framework to describe the localisation of all proteins in the mammalian cell.

Professor Mike Fellows was awarded a Humboldt Fellowship by the Alexander von Humboldt-Stiftung. He will spend part of 2006-2007 at the Institute of Advanced Study at Durham University (UK).

Professor Geoff McLachlan was awarded an ARC Professorial Fellowship in support of his research on multivariate methods for analysis of microarray gene expression data.

Dr Sean Grimmond was awarded a Senior Research Fellowship by the National Health and Medical Research Council.

ACB Advisory Board member Professor John Mattick received the 2006 CSIRO Eureka Prize for Leadership in Science in recognition of his "scientific leadership, organisational vision and exemplary commitment in the establishment and development of the Australian Genome Research Facility and the Institute for Molecular Bioscience".

ACB Advisory Board member Dr Marc Wilkins helped organise the 2006 Bioinformatics Australia meeting in Sydney, held in association with the annual meeting of AusBiotech. A/Professor Phoebe Chen and Dr Rohan Teasdale are members of the BA advisory board.

At Australian Computer Science Week in Hobart, CORE (Computing Research and Education Association of Australia) presented top dissertation awards to two postgraduates associated with ACB. Dr Nic Geard (University of Queensland) was joint 2006 winner of the Australasian Distinguished Doctoral Dissertation Award for his thesis "Artificial ontogenies: a computational model of the control and evolution of development." Dr Elena Prieto Rodriguez of Newcastle Bioinformatics Initiative won the 2006 Australasian Distinguished Doctoral Dissertation Special Commendation marking second place nationally for her 2005 thesis "Systematic kernelisation in FPT algorithm design".

1. Research findings

Quality of publications

ACB researchers continue to perform very well in peer-reviewed publication, indicating the innovative and forefront nature of our research. In 2006 we published 84 C1 papers, a more than twofold increase over 2005. This sharp increase graphically illustrates the impetus that ARC funding of ACB (which became available to us in late 2004) has given bioinformatics research in Australia. Of these 84 papers, 46 (55%) appeared in journals with 2005 impact factor ≥ 3.0 , 32 (38%) with $IF \geq 5.0$, and 11 (13%) with $IF \geq 10.0$. Our impact beyond bioinformatics *per se* is reflected by numerous publications in highly prestigious journals including *Blood*, *Cell*, *Genome Biology* (4), *Genome Research* (2), the *Nature* journals (4), *PLoS Genetics* (5), *Proceedings of the National Academy of Sciences USA* (2), *Systematic Biology*, and *Trends in Genetics*. Our work also appeared in top bioinformatics journals including *Bioinformatics* (3, with a fifth in e-publication ahead of print), *BMC Bioinformatics* (9), and *PLoS Computational Biology*.



Number of publications

The table below summarises the published output by our investigators and research groups that appeared with a 2006 publication date. We do not count papers that at the end of December 2006 were available electronically in advance of print publication, or remained in press, under review, submitted or in preparation. Nor do we count publications by researchers associated with, but not technically part of, our Centre. For details please refer to Appendix 1: Publications and Papers.

Category	Quantity
A3 Books – edited	2
B Book chapter	7
C1 Journal articles – articles in scholarly refereed journal	84
C2 Journal articles – other contribution to refereed journal	2
C3 Journal articles – non-refereed articles	5
E1 Conference – full written paper – refereed proceedings	29
E2 Conference – full written paper – non-refereed proceedings	2
E4i Conference – edited volume of conference proceedings	2
E4 Conference – unpublished presentations	5
G Computer software	5
Other academic outputs	
– posters	2
– reports	2
– theses	7
TOTAL	154

Patents and trademarks

No patents were awarded to our investigators during 2006.

In September and October 2006, in conjunction with IMBcom we filed for registration of Visible Cell™ as an international trademark in Australia, New Zealand, USA, Europe and the United Kingdom. We own relevant Web domain names including www.visiblecell.org and www.visiblecell.com.

Development of the Visible Cell™ is proceeding in three successive stages: (1) implementation of an initial prototype integrated with scalable data infrastructure; (2) development of subsequent prototypes with enhanced functionality to address real-life biological problems; and (3) release and support of stable, highly capable versions for production-scale use, with periodic updates driven by user demand. With IMBcom we are developing a commercial partnering strategy for medium- and longer-term development.

IMBcom Pty Ltd, a technology management corporation owned by The University of Queensland, conducted BioBusiness training sessions in May (introductory workshop) and July (three-day retreat) 2006. The introductory workshop provided first-year students an overview of the concepts of bio-business. The second was a three-day retreat for third-year students, covering issues in commercialisation as well as career options. Several months prior to the retreat, students were assigned to groups and prepared, with the help of an IMBcom mentor, a business plan for presentation at the retreat. The two BioBusiness sessions form a compulsory component of the postgraduate programme for IMB students. ACB sponsors the participation of ACB students from universities other than UQ, and from other faculties within UQ. ACB's postdoctoral fellow and students attending the introduction session in 2006 were:

- Denis Bauer (Institute for Molecular Bioscience, UQ)
- Karin Kassahn (Institute for Molecular Bioscience, UQ)
- Emma Redhead (Institute for Molecular Bioscience, UQ)
- Michael Piechota (Institute for Molecular Bioscience, UQ)
- Chang Jin Shin (Institute for Molecular Bioscience, UQ)

Invitations to address and participate in conferences

International conferences

Chief Investigators

In 2006, ACB investigators and members of their research groups accepted 115 invitations to speak at, or otherwise contribute significantly to, national and international scientific conferences. Participation in ACB's Winter School is not counted in this section. We list relevant invitations for ACB associates Dr Brad Marsh and Dr Pablo Moscato, but do not count them toward our KPI.

Dr Timothy Bailey

- Program Committee, Intelligent Systems for Molecular Biology (ISMB2006), Fortaleza (August)

Prof. Kevin Burrage

- Sanum Conference, Stellenbosch (April)
- Positive Systems: Theory and Applications (Posta06) Conference, Grenoble (August)

A/Prof. Phoebe Chen

- Steering Committee Chair, 12th International Multimedia Modelling Conference (MMM2006), Beijing (January)
- Eighth Asia Pacific Web Conference, Harbin (January)
- Conference Co-chair and Steering Committee Chair, Fourth Asia-Pacific Bioinformatics Conference (APBC2006), Taipei (February)
- Workshop on Data Mining for Biomedical Applications (BioDM2006) in conjunction with 10th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2006), Singapore (April)
- Fourth International Conference on Active Media Technology (AMT 2006), Brisbane (June)
- IEEE International Conference on Multimedia & Expo (ICME2006), Toronto (July)
- Session Chair, 18th International Conference on Pattern Recognition (ICPR2006), Hong Kong (August)
- Publicity Co-Chair, 2006 Workshop on Pattern Recognition in Bioinformatics (PRIB'06), Hong Kong (August)

- 6th Workshop on Algorithms in Bioinformatics (WABI2006), Swiss Federal Institute of Technology, Zürich (September)
- 10th International Conference on Knowledge-Based & Intelligent Information & Engineering Systems (KES2006), Bournemouth (October)
- 6th IEEE Symposium on Bioinformatics and Bioengineering (BIBE-06), Washington DC (October)
- 17th International Conference on Genome Informatics (GIW2006), Yokohama (December)
- 2006 International Workshop on Medical and Bio Informatics (in conjunction with ICS 2006), Taipei (December)

Prof. Michael Fellows

- ACM Symposium on Theory of Computing (STOC 2006), Seattle (May)
- SIAM Discrete Mathematics Conference (SIAM-DM 2006), Victoria, Canada (June)
- Keynote, 2nd International Workshop on Parameterized and Exact Computation (IWPEC2006), Zürich (September)
- European Symposium on Algorithm (ESA 2006), Zürich (September)
- Second Algorithms and Complexity in Durham Workshop (ACiD 2006), Durham (September)
- Keynote, International Conference on Discrete Mathematics (ICDM 2006), Bangalore (December)

Dr Markus Hegland

- Conference on High Performance Scientific Computing, Hanoi (March)

Prof. Geoffrey McLachlan

- One-day workshop at The University of Waikato, Hamilton, NZ (April)
- Australian Statistical Conference/ New Zealand Statistical Association Conference 2006, Auckland (July)
- XXIIIrd Meeting of International Biometric Society, Montréal (July)
- One-day workshop at the Joint Statistical Meetings 2006, Seattle (August)
- Compstat 2006, Rome (August-September)
- Knowledge Extraction and Modelling IASC-INTERFACE-IFCS Workshop, Isle of Capri (September)

Dr Brad Marsh

- Congress ALFEDIAM 2005 (l'Association de Langue Française pour l'Etude du Diabète et des Maladies Métaboliques), Paris (March)
- Microscience 2006 - Royal Microscopy Society, London (June)
- 4th International Congress on Electron Tomography, San Diego (November)

A/Prof. Pablo Moscato

- Program Committee, 6th European Conference on Evolutionary Computation in Combinatorial Optimization, Budapest (April)
- Program Committee, 4th European Workshop on Evolutionary Computation and Machine Learning in Bioinformatics, Budapest (April)
- Program Committee, IEEE Congress on Evolutionary Computation (Special Session on Evolutionary Clustering), Vancouver (July)
- Program Committee, Ninth International Conference on Parallel Problem Solving from Nature, Reykjavic (September)
- Program Committee, International Conference on Hybrid Information Technology, Jeju Island (November)

Prof. Bernard Pailthorpe

- Host and Co-Chair, Pacific Rim Applications and Grid Middleware Assembly Workshop, Townsville (March)
- Host and Co-Chair, Coral Reef Ecological Observatory Network / Great Lakes Ecological Observatory Network (CREON/GLEON) Conference, Townsville (March)

Prof. Mark Ragan

- Keynote, Fourth Asia-Pacific Bioinformatics Conference (APBC) 2006, Taipei (February)
- Korea Society for Biochemistry and Molecular Biology Annual Meeting 2006, Seoul (May)
- 3rd International *E. coli* Alliance Conference on Systems Biology (IECA 2006), Jeju Island (October)
- International Symposium on Bioinformatics Education and Research, Yokohama (December)
- 17th International Conference on Genome Informatics (GIW2006), Yokohama (December)

Dr Isidore Rigoutsos

- Plenary presentation, Keystone Symposium on Signaling Networks, Vancouver, British Columbia, Canada (January)
- Plenary presentation, International Congress on the Applications of Applied Mathematics (SIAM), Santiago, Chile (March)
- Keynote, 5th International BioExpo, Tokyo (May)
- Plenary presentation, 4th International RNAi-Boston Meeting, Waltham, MA (May)
- Program Committee, 3rd Conference on Email and Anti-Spam (CEAS 2006), Mountain View, CA (July)
- Program Committee, 5th IEEE Computational Systems Bioinformatics Conference (CSB 2006), Stanford, CA (August)
- Program Committee, 14th Annual Conference on Intelligent Systems for Molecular Biology (ISMB '06), Fortaleza, Brasil (August)
- Program Committee, 5th European Conference on Computational Biology (ECCB 2006), Eilat, Israel (September)

Dr Rohan Teasdale

- Expression Cluster Workshop, RIKEN Institute, Yokohama (December)
- 46th Annual Meeting of The American Society of Cell Biology, San Diego (December)

Prof. Janet Wiles

- Program Committee, Tenth International Conference on the Simulation and Synthesis of Living Systems, Bloomington, USA

Prof. Xiaofang Zhou

- Program Committee Co-chair, 8th Asia Pacific Web Conference (APWeb 2006), Harbin (January)
- IASTED International Conference on Database and Applications (DBA 2006), Innsbruck (February)
- 10th International Conference on Extending DataBase Technology (EDBT 2006), Munich (March)
- 20th IEEE Advanced Information Networking and Applications (AINA 2006), Vienna (April)



- 11th International Conference on Database Systems for Advanced Applications (DASFAA 2006), Singapore (April)
- 22nd International Conference on Data Engineering (ICDE 2006), Atlanta (April)
- Workshop on Mobile Location-Aware Sensor Networks (in conjunction with MDM 2006), Nara (May)
- The 15th International Conference on World Wide Web (WWW'06), Edinburgh (May)
- International Conference on Advanced Visual Interfaces (AVI 2006), Venice (May)
- 7th International Conference on Web-Age Information Management (WAIM 2006), Hong Kong (June)
- 26th IEEE International Conference on Distributed Computing Systems (ICDCS 2006), Lisboa (July)
- 6th International Conference on Web Engineering (ICWE'06), Menlo Park, California (July)
- 2006 IFIP International Conference on Embedded and Ubiquitous Computing (EUC 2006), Seoul (August)
- First International Conference on Knowledge Science, Engineering and Management (KSEM 2006), Guilin (August)
- 17th International Workshop on Database and Expert Systems Applications (DEXA 2006) Workshop on Bioinformatics, Krakow (September)
- 1st International Workshop on XML Data Management Tools & Techniques (XANTEC'06), in conjunction with Database and Expert Systems Applications (DEXA 2006), Krakow (September)
- 4th International Workshop on Biological Data Management (BIDM'06), in conjunction with DEXA 2006, Krakow (September)
- IBM Dublin CASCON 2006 Symposium, in association with 16th Annual International Conference on Computer Science and Software Engineering (CASCON 2006), Mulhuddart, Ireland (October)
- 13th International Conference on Management of Data (COMAD 2006), New Delhi (December)
- 2006 Workshop on Mining Spatial and Spatiotemporal Data (SSTDM), in conjunction with IEEE International Conference on Data Mining (ICDM), Hong Kong (December)

- IEEE ICDM 2006 Workshop on Data Mining in Bioinformatics (DMB 2006), Hong Kong (December)

Postdoctoral fellows, Research Officers and students

Dr Jiyuan An

- 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI-06), Hong Kong (December)

Dr Robert Beiko

- Program Committee, Fourth Annual RECOMB Satellite Meeting on Comparative Genomics, Montréal (September)

Dr Aaron Darling

- 3rd International *E. coli* Alliance Conference on Systems Biology (IECA 2006), Jeju Island (October)

Dr Martin Frith

- 14th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB 2006), Fortaleza (August)
- 17th International Conference on Genome Informatics (GIW2006), Yokohama (December)

Dr Jim Hanan

- 3rd International Conference on Legume Genomics and Genetics, Brisbane (April)
- Session Coordinator and Chair, 8th International Congress on Plant Molecular Biology, International Society for Plant Molecular Biology, Adelaide (August)

Ms Helen Huang

- 17th International Workshop on Database and Expert Systems Applications (DEXA 2006) Workshop on Bioinformatics, Krakow (September)

Ms Ann Morrison

- ACM Multimedia 2006, Santa Barbara, California (October)

In addition, eight ACB postgraduate students participated in our joint student workshop with Bioinformatics Institute New Zealand (Auckland, 11–15 July 2006). For details, please see page 49.

National conferences

Chief Investigators

Dr Timothy Bailey

- Organiser, Workshop for Intelligent Systems in Bioinformatics (WISB 2006), Hobart (December)

A/Prof. Phoebe Chen

- Bioinformatics Australia 2006 "Connecting Australian Bioinformatics", Sydney (November)
- 19th Australian Joint Conference of Artificial Intelligence 2006 (AI 2006), Hobart (December)

Dr Sean Grimmond

- ComBio 2006, Australian Society for Biochemistry and Molecular Biology, Brisbane (September)

Dr Markus Hegland

- Scientific Committee member, ANZIAM Annual Meeting, Mansfield (February)
- Australian Mathematical Society 50th Annual Meeting, Sydney (September)

Prof. Geoffrey McLachlan

- BioInfoSummer 2006 – ICE-EM Summer Symposium in Bioinformatics, Canberra (December)
- Workshop in Intelligent Systems in Bioinformatics, Hobart (December)

Dr Brad Marsh

- Australian Conference on Microscopy & Microanalysis, Sydney (February)

A/Prof. Pablo Moscato

- 28th Annual Scientific Meeting of the ANZ Breast Cancer Trials Group (ANZBCTG), Cairns (July)
- Hunter Medical Research Institute (HMRI) Conference on Translational Cancer Research, Newcastle (September)

Dr Ralf Muhlberger

- OzCHI Virtual Communities Workshop, Sydney (November)

Prof. Bernard Pailthorpe

- Computing the Future Symposium [50th Anniversary of *SILIAC*, the first computer in an Australian university], Sydney (September)
- FEAST Workshop (European Union Seventh Framework Program), Canberra (November)

Prof. Mark Ragan

- 2006 e-Health Research Colloquium, Brisbane (March)
- University of Queensland eResearch Workshop, Brisbane (June)
- Inaugural Australian Materials Informatics Workshop, Brisbane (November)
- Program Committee Chair, Workshop on Intelligent Systems for Bioinformatics, Hobart (December)

Dr Rohan Teasdale

- ComBio 2006, Australian Society for Biochemistry and Molecular Biology, Brisbane (September)
- Systems Biology and Bioinformatics Symposium Stream Coordinator, ComBio 2006, Brisbane (September)

Prof. Janet Wiles

- Keynote, CSIRO Complex Systems Science, Sydney (August)

Prof. Xiaofang Zhou

- 2006 Workshop on Middleware for Sensor Networks 2006, in conjunction with 7th ACM/IFIP/USENIX Middleware Conference, Melbourne (November)

Postdoctoral fellows, Research Officers and students

Dr Richard Bean

- BioInfoSummer 2006 – ICE-EM Summer Symposium in Bioinformatics, ANU, Canberra (December)

Dr Lynn Fink

- ComBio 2006, Australian Society for Biochemistry and Molecular Biology, Brisbane (September)

Dr Martin Frith

- WEHI Bioinformatics Seminar, Melbourne (October)

Dr Jim Hanan

- ComBio 2006, Australian Society for Biochemistry and Molecular Biology, Brisbane (September)
- Program Committee, Workshop on Intelligent Systems for Bioinformatics held in conjunction with the Australian Joint Conference on Artificial Intelligence, Hobart (December)

Dr Nicholas Hamilton

- ComBio 2006, Australian Society for Biochemistry and Molecular Biology, Brisbane (September)

- Program Committee, Workshop on Intelligent Systems for Bioinformatics in held in conjunction with the Australian Joint Conference on Artificial Intelligence, Hobart (December)

Mr Steven Livingstone

- HCSNet Workshop on the Science of Music Perception, Performance and Cognition, Sydney (September)
- HCSNet Workshop on Human and Machine Speech, Sydney (September)
- HCSNet Workshop on Perception and Action, Sydney (September)

Invitations to visit leading international research groups and research centres

Chief Investigators

Dr Timothy Bailey

- RIKEN Genome Sciences, Yokohama (January)
- Computational Biology Research Centre, Tokyo (January)
- Human Genome Center, The Institute of Medical Science, The University of Tokyo University (January)
- San Diego Supercomputer Center, USA (September)
- Department of Genome Sciences, and Department of Computer Science and Engineering, University of Washington, Seattle (September)

Prof. Kevin Burrage

- Mathematics Department, Università degli Studi di Napoli Federico II (September)
- Mathematics Department, Università degli Studi di Roma La Sapienza (September)
- Mathematics Department, Università degli Studi di Urbino Carlo Bo (September)
- Bioinformatics Institute, University of Auckland (July)

A/Prof. Phoebe Chen

- Department of Computer Science, Capital Normal University, Beijing, (January)
- Academy of Mathematics and System Science, Chinese Academy Science, Beijing, (January)
- Graduate Institute of Clinical Medicine, Chang Gung Memorial Hospital (Kaohsiung), Kaohsiung (February)

- Department of Surgery and Chang Gung Bioinformatics Centre, Chang Gung Memorial Hospital (Lin Kao), Taipei (February)
- Departments of Biochemistry & Molecular and Medical Genetics, University of Toronto (July)
- School of Computer Science, University of Waterloo, Canada (July)
- Department of Statistics (Bioinformatics Group) and Department of Computer Science, Oxford University (October)

Prof. Michael Fellows

- Parameterized Complexity Research Group, University of Freiburg (September)
- Algorithms Research Group, ETH Zürich (September)
- Complexity Theory Research Group, University of Maryland (December)

Dr Markus Hegland

- Department of Mathematics, National University of Singapore (April)

Dr Brad Marsh

- Max Planck Institute of Biochemistry, Martinsried (March)
- Centre for Molecular Medicine and Therapeutics, University of British Columbia (April)
- Cell Biology and Metabolism Branch, National Institute of Child Health & Human Development (NICHD), NIH, Bethesda (April)
- Department of Molecular Cell Biology, Leiden University Medical Center (June)
- FEI Company, Eindhoven, The Netherlands (July)
- National Resource for Automated Molecular Microscopy, Center for Integrative Molecular Biosciences, Scripps Research Institute, La Jolla (November)

A/Prof. Pablo Moscato

- Departamento de Lenguajes y Ciencias de la Computación, Universidad de Málaga, Spain (November)
- Departament de Llenguatges i Sistemes Informatics, Universitat Politècnica de Catalunya, Spain (UPC) (December)

Prof. Bernard Pailthorpe

- California Institute for Telecommunications and Information Technology (CallT), San Diego (February)



Dr Isidore Rigoutsos

- Medical School, Thomas Jefferson University, Philadelphia (January)
- Plenary presentation, University of Virginia, Charlottesville, Virginia (April)
- Colloquium participant, Genome Institute of Singapore (GIS), Singapore (May)
- Colloquium participant, Japan Biological Information Research Center (JBIRC), Tokyo (May)
- Department of Biology, University of Tokyo, Tokyo (May)
- Department of Microbiology, Mount Sinai Medical School, New York (May)
- Department of Genetics, Cambridge University, Cambridge, UK (October)
- BIOMAPS Institute for Quantitative Biology, Rutgers University (October)
- Plenary presentation, 1st International microRNAs Europe Meeting, University of Cambridge, UK (November)
- CBR Institute for Biomedical Research, Harvard Medical School, Harvard, MA. (December)

Prof. Xiaofang Zhou

- Multimedia Information Networking Lab, Tamkang University, Taiwan (May)
- National Key Centre in Software Engineering, Wuhan University, China (May)
- School of Computer Science and Technology and ChinaGrid Project, Huazhong University of Science and Technology, China (May)
- Department of Computer Science, City University of Hong Kong (June)
- SAP Laboratories, China (September)
- Department of Social Informatics Graduate School of Informatics, Kyoto University, Japan (December)
- Department of Computer Science, Fudan University, China (December)
- School of Information, Renmin University, China (December)

Postdoctoral fellows, Research Officers and students

Dr Robert Beiko

- Department of Biology, Memorial University of Newfoundland, Canada (February)
- Department of Computer Science, Carleton University, Canada (April)
- Faculty of Computer Science, Dalhousie University, Canada (May)

Dr Jim Hanan

- External assessor on PhD thesis, Crop and Weed Ecology Group, Wageningen University (March)

Commentaries about the Centre's achievements

ACB was featured in a special issue of *Asia-Pacific Biotech News* (volume 10, issue 24: 30 December 2006) on "Bioinformatics in Asia Pacific".

Prof Mark Ragan was interviewed by Mr Bernard Lane, *The Australian* (8 March 2006) "Direct to the public – A new online biology journal of peer review promises to offer a lively forum for discussion".

Dr Rohan Teasdale's LOCATE database was featured in a news item "GPS for proteins" in *Science* (14 July 2006, <http://www.sciencemag.org/netwatch/>). LOCATE was also highlighted in Signal Transduction Knowledge Environment (STKE) and the Canadian Bioinformatics Network online newsletter.

A report on The First Australia-Taiwan Bioinformatics Symposium (see our 2005 Annual Report), which we co-organised with the Institute of Bioinformatics and Structural Biology, National Tsing Hua University, Hsinchu (6–9 September 2005), is available online at <http://www.flickr.com/photos/changturtle/68319754/>.

Evidence of interdisciplinary linkages within the Centre

Much of ACB's research is conducted by interdisciplinary teams within which many of our postgraduate and postdoctoral researchers are co-supervised or mentored by researchers from different fields, *e.g.* bioscience and maths/IT. In this way we help to prepare the next generation of researchers to recognise and appreciate problems in systems biology, to think laterally into other fields and leverage insights and methods from multiple disciplines, and to work in and lead multi-disciplinary teams.

Of our 16 projects and sub-projects active during 2006, 13 (81%) included personnel with advanced degrees in different disciplines; 22 of our 32 PhD students (69%) and 18 of our 19 PDFs (95%) worked in these 13 projects. Of our 28 papers and posters during 2006 with at least two Centre investigators or students as co-authors, 17 (61%) of those co-authors hold advanced degrees in different disciplines. 11 of 21 (55%) co-authored C1 journal publications were interdisciplinary in this sense. Our All-Hands Meeting and Winter School are intentionally interdisciplinary, and all ACB students and postdoctoral researchers are strongly encouraged to attend.

Three ACB projects (2.4, 2.5 and 3.2.2) are collaborative with the ARC Centre for Complex Systems (Dr Jim Hanan) or other complex systems researchers (Prof. Janet Wiles).

ACB on the Web

Our Centre's website, bioinformatics.org.au, serves as a virtual hub for the Australian bioinformatics community. It provides a consistent medium that facilitates communication about conferences, workshops and job opportunities. One of its more popular features is online access to digitally recorded video lectures from our annual Winter School in Mathematical and Computational Biology.

A key function of ACB's website is the hosting of computational tools and datasets. The website provides a home for tools that might not otherwise be easily available to the research community. New tools produced by ACB members published and hosted on our website this year include the programs EEP (Beiko and Hamilton, *BMC Evolutionary Biology*) and GONOME (Stanley, Bailey and Mattick, *BMC Bioinformatics*); the databases LOCATE (Fink *et al.*, *Nucleic Acids Research*) and PhosphoregDB (Forrest *et al.*, *BMC Bioinformatics*); and the dataset LGT144 (Beiko, Harlow and Ragan, *Proceedings of the National Academy of Sciences USA*).

Our ACB website is visited by over 500 unique users a month from all over the world – underscoring that Australian bioinformatics garners international interest in addition to supporting local research.

<http://www.bioinformatics.org.au/>



LOCATE

One of the biological resources that ACB hosts is LOCATE, a curated database that houses data describing the membrane organisation and subcellular localisation of proteins from the RIKEN mouse and human protein sequence sets. The data contained in LOCATE are the direct result of several ACB publications (Davis *et al.*, *In Silico Biology*; Davis *et al.*, *PLoS Genetics*; and Aturaliya *et al.*, *Traffic*). Protein membrane organisation is predicted by the high-throughput computational pipeline MemO, and the subcellular locations were determined by a high-throughput, immunofluorescence-based assay and by manually reviewing peer-reviewed publications.

LOCATE, described in the annual *Nucleic Acids Research Database Issue* (Fink *et al.*), represents the first effort to catalogue the experimentally verified subcellular location and membrane organisation of mammalian proteins using a high-throughput approach. It provides localisation data for about 40% of the mouse and human proteomes, and can be accessed *via* the main website, or directly at locate.imb.uq.edu.au.

LOCATE was recently profiled in Science's NetWatch column (July 2006) and has received over 1000 unique visitors in a single month. LOCATE was a highly collaborative effort that involved ACB members Lynn Fink, Kelly Hanson, and Rohan Teasdale and several local collaborators.

<http://locate.imb.uq.edu.au/>



The Visible Cell™

The Visible Cell™ website, launched in mid-October, details the aims and expected outcomes of the Visible Cell™ project. The site provides background information on the researchers and developers involved, and gives details on individual projects that contribute to the Visible Cell™ effort. The project team can be contacted *via* email from the site. In the near future the Cell Visualiser, a core component of the Visible Cell™, will be made available for download *via*. The site can be accessed either *via* our Centre's main site, or at www.visiblecell.com.

<http://www.visiblecell.com/>



2. Research training and professional education

A ground-truth indication of the value of our training environment comes from the positions our postgraduates and postdocs have taken up after ACB. So far these include employment with the Australian Centre for Plant Functional Genomics; a Tier II Canada Research Chair; Computational Biology Research Centre, National Institute of Advanced Industrial Science (Japan); IBM Healthcare & Life Sciences (USA); the International Centre for Tropical Agriculture; Oracle (Australia); and San Diego Supercomputer Center. In mid-2006 Dr Lindsay Hood, Visible Cell™ project manager, moved to the Australian Partnership for Advanced Computing to lead the APAC Grid program.

Recruitment

We recruited the following highly qualified staff in 2006:

- Application Developer Mr Tim McComb (to start in January 2007) from the School of Information Technology & Electrical Engineering, UQ into Project 3.2.1 (Concept and engineering of the Visible Cell™).
- Database Administrator/Developer Mr David Wood from Australian Genome Research Foundation, into Project 3.1.1 (Data environment for the Visible Cell™).
- Postdoctoral Fellow Dr Aaron Darling from the University of Wisconsin, into Project 2.6 (Ortholog mapping). Aaron was awarded one of the inaugural NSF International Bioinformatics traineeships, to work for two years with Prof. Mark Ragan (2006-2008) followed by one year with Dr Jonathan Eisen (University of California, Davis).
- Postdoctoral Fellow Dr Karin Kassahn from James Cook University, into Project 2.6 (Ortholog mapping).
- Research Assistant Mr Geoffrey Faulkner from the University of Queensland (part-time during 2006), into Project 1.1 (Transcriptional complexity of mammalian cells).
- Research Assistant Mr Timothy Sullivan from Queensland University of Technology, into Projects 3.1.1 (Concept and engineering of the Visible Cell™) and 3.1.2 (Integrating database technologies and visual analysis).

- Research Fellow Dr Fran Rosamond from The University of Newcastle (part-time during 2006), into Project 2.1 (Hybrid algorithms for large-scale network and graph optimisation).
- Research Officer Ms Melissa Brown from Griffith University (50% during 2006), into Project 1.1 (Transcriptional complexity of mammalian cells).
- Research Officer Mr Alistair Forrest from the Institute for Molecular Bioscience, into Project 1.1 (Transcriptional complexity of mammalian cells).
- Research Officer Ms Kelly Hanson via promotion within ACB, into Project 1.3 (Generation of the experimental data for modelling cellular compartments).
- Research Officer Dr Muhammad Shoaib Sehgal from Monash University (to start in March 2007), to work at the interface between Projects 3.1.2 (Integrating database technologies and visual analysis) and 3.2.2 (Spatial modelling of cellular systems).
- Research Officer Dr James Watson from the School of Information Technology & Electrical Engineering, UQ, joint appointment with ARC Centre for Complex Systems, into Project 2.5 (Complex systems modelling of regulatory and protein-protein interaction networks).

2006 Winter School in Mathematical and Computational Biology (26–30 June 2006)

Our Winter School was once again the main advanced training event in our year. The 2006 Winter School featured 9 international, 12 national and two commercial speakers. We welcomed registrants from 41 distinct institutions, mostly from Australia and New Zealand. Approximately half of the participants were postgraduates and advanced undergraduates; one-third have a biology background, with most others coming from mathematics, statistics, computer science and IT. Among the Australian registrants, 60% represented Queensland universities. Again we received highly positive written feedback from the participants about the topics and quality of presentations.

We thank the ARC Research Network in Enterprise Information Infrastructure (EII), Queensland Parallel Supercomputing Foundation (now Queensland Cyber Infrastructure Foundation), The MathWorks Australia, and the Institute for Molecular Bioscience for generous sponsorship.

Each of the five days focused on a different theme:

Day 1

Machine learning applied to sequence analysis

Prof. Mark Borodovsky
School of Biology
Georgia Institute of Technology, USA

Dr Tim Bailey
ARC Centre in Bioinformatics, and
Institute for Molecular Bioscience
The University of Queensland

Dr Paul Horton
Computational Biology Research Centre
National Institute of Advanced Industrial
Science and Technology, Japan

Dr Martin Frith
ARC Centre in Bioinformatics, and
Institute for Molecular Bioscience
The University of Queensland

Day 2

Combinatorial optimisation in bioinformatics

Dr Regina Berretta
Newcastle Bioinformatics Initiative
The University of Newcastle

A/Prof. Pablo Moscato
Director
Newcastle Bioinformatics Initiative
The University of Newcastle

Dr Tim Bailey
ARC Centre in Bioinformatics, and
Institute for Molecular Bioscience
The University of Queensland

Mr Paul Taylor
The MathWorks Australia

Dr James J. Cai
Department of Biological Sciences
Stanford University, USA

A/Prof. Andrew Roger
(evening public lecture)
Canadian Institute for Advanced Research,
and Department of Biochemistry and
Molecular Biology, Dalhousie University,
Canada

Day 3

Evolutionary genomics

Professor Mike Hendy
Director
Allan Wilson Centre for Molecular Evolution
Institute of Fundamental Sciences
Massey University, NZ

Dr Barbara Holland
Allan Wilson Centre for Molecular Evolution
Institute of Fundamental Sciences
Massey University, NZ

Dr Michael Charleston
School of Information Technologies
The University of Sydney

Dr Charles Semple
Department of Mathematics and Statistics
The University of Canterbury, NZ

Dr Alexei Drummond
Bioinformatics Institute New Zealand
Department of Computer Science
The University of Auckland, NZ

Day 4

Microarrays and analysis of microarray data

Dr Sean Grimmond
ARC Centre in Bioinformatics, and
Institute for Molecular Bioscience
The University of Queensland

Professor Terry Speed
Department of Statistics
University of California Berkeley, USA;
and Walter and Eliza Hall Institutes of
Medical Research

Dr Liat Ben-Tovim Jones
ARC Centre in Bioinformatics, and
Institute for Molecular Bioscience
The University of Queensland

Dr Toni Reverter
Bioinformatics Group
CSIRO Livestock Industries

Professor Kim-Anh Do
Department of Biostatistics and
Applied Mathematics
M.D. Anderson Cancer Center
The University of Texas, USA

Professor Geoff McLachlan
ARC Centre in Bioinformatics, and
Department of Mathematics
The University of Queensland

Day 5

The e-research / Grid paradigm

Professor David Abramson
Faculty of Information Technology
Monash University

Professor Jane Hunter
School of Information Technology &
Electrical Engineering
The University of Queensland

Professor Ah Chung Tsoi
Director
Monash e-Research Centre
Monash University

Dr David Hansen
e-Health Research Centre
CSIRO

A/Prof. Dr Tin Wee Tan
Yong Loo Lin School of Medicine
National University of Singapore

ACB-BINZ Bioinformatics Student Symposium (New Zealand)

ACB sponsored the airfares of eight ACB postgraduate students and four ACB Chief Investigators/associates to attend a student symposium hosted by Bioinformatics Institute New Zealand (Auckland, 11-15 July 2006). The program included research presentations by students and established researchers, as well as social events. All students were required to make a 15-minute presentation on their thesis work.

ACB investigators and associates:

- Dr Regina Berretta
- Dr Mikael Bodén
- Prof. Kevin Burrage
- A/Prof. Pablo Moscato

ACB postgraduate students:

- Cheong Xin Chan (UQ)
- Wagner Emanuel Costa (Newcastle)
- John Hawkins (UQ)
- Keith Knapp (Deakin)
- Shev MacNamara (UQ)
- Mario Inostroza Ponta (Newcastle)
- Chang Jin Shin (UQ)
- Chin Foon Khoo (ANU)

International exchange program

Mr Leo Huber, a MSc student from ETH Zürich, spent much of 2006 with Dr Markus Hegland at our ANU node, applying a new adaptive sparse grid technique to solve the Chemical Master Equation for large (100-protein) systems.

Other research training and professional education

In 2006, ACB researchers were involved in research training and professional education other than through our Winter School:

Prof. Kevin Burrage

- ACB-Bioinformatics Institute (NZ) Student Symposium, Auckland (July)
- 2006 School in Computational Cell Biology, Università di Modena (September)

Prof. Geoff McLachlan

- American Statistical Association workshop on Methods and Computational Tools for the Screening and Classification of Microarray Gene Expression Data, given as part of the ASA Continuing Education Program, Seattle (August)
- Computational Statistics International School, Università di Napoli (September)

Dr Brad Marsh

- Seminar, Garvan Institute, Sydney (May)
- SVI Seminar Series, St. Vincent's Institute, Melbourne (September)
- Special Seminars, Walter and Eliza Hall Institute of Medical Research, Melbourne (September)
- Seminar, Advanced Computational Modelling Centre, UQ (November)

Prof. Bernard Pailthorpe

- Host, Storage Resource Broker training workshop, Brisbane (March)
- APAC Summer School on Computational Science, Brisbane (December)



Undergraduate courses in the Centre's area

Our investigators and researchers participated in the following for-credit courses and programs in 2006:

Dr Timothy Bailey

- BIOL3014 Advanced Bioinformatics (coordinator and lecturer)

A/Prof. Phoebe Chen

- Bachelor of Information Technology (Multimedia Technology) (course coordinator)
- SIT781 Introduction to Bioinformatics (unit chair and lecturer)

Dr Sean Grimmond

- BIOL3004 Genomics and Bioinformatics (lecturer)
- BIOT7009 Emerging Technologies (lecturer)

Dr Jim Hanan

- BOTN2002 Functional Plant Biology (guest lecturer)
- Math3104 Mathematical Biology (guest lecturer)

Dr Nicholas Hamilton

- BIOL3044 Vacation Research Project (supervisor)

Dr Brad Marsh

- BIOL3006 Molecular Cell Biology (lecturer)
- BIOC3004 Structural Biology: Macromolecules (lecturer)

Prof. Bernard Pailthorpe

- MATH3201 Advanced Computing (organiser and lecturer)

Prof. Mark Ragan

- BIOL3004 Genomics and Bioinformatics (introductory lecture)

Dr Rohan Teasdale

- BIOL3004 Genomics and Bioinformatics (lecturer)

Senior high school student research scheme

PhD student Mr Cheong Xin Chan, through the CSIRO Student Research Scheme 2006, supervised two high school students (Callan Cavaghan from Grace Lutheran College and Liam Nolan from Marist College) on a 2-week mini project, "Finding the odds in DNA sequences".

3. International, national and regional links and networks

Our researchers have strong research linkages with their counterparts in many institutions. ACB plays an active role in international, national and regional bioinformatics initiatives in support of our mission.

ACB has been instrumental in catalysing the identification of bioinformatics as an institutional research priority at several of our partner institutions. In 2006 the Newcastle Bioinformatics Initiative, an ACB strategic partner since 2004, became a founding partner in the new Priority Research Centre for Bioinformatics, Biomarker Discovery and Information-Based Medicine at The University of Newcastle.

ACB was proactive in the start-up of Queensland Facility for Advanced Bioinformatics (QFAB), launched on 21 December 2006. QFAB, a joint initiative of seven institutions, received strategic support from the NIRAP program of Queensland Government (see page 40). Several ACB investigators were active in state and national initiatives as part of the NCRIS process.

In October 2006, Bioinformatics Australia named ACB postdoctoral fellow Dr Lucia Santoso, from our ANU node, to the role of Executive Officer of the new Australian Bioinformatics Network. At the annual Bioinformatics Australia meeting (Sydney, 21–22 November 2006), Lucia organised interactive discussions on bioinformatics research, development and infrastructure in Australia.

Prof. Mark Ragan is a member of the Board of Management of the DART e-Research Project. DART is headquartered at Monash University with significant nodes at James Cook University and UQ.

Prof Mark Ragan served on the programme committee for GCCB 2006 (International Workshop on Distributed, High-Performance and Grid Computing in Computational Biology, 10-13 September 2006 in Eilat, Israel. This workshop was postponed to January 2007.

Internationally, the unique profile of expertise within ACB attracted commercial partnerships with Applied Biosystems Inc. (Foster City, California) and Pfizer Research Technology Center (Cambridge, Massachusetts). We continue involvement in the FANTOM consortia organised by RIKEN Genome Sciences (Japan); the Asia Pacific Bioinformatics Conference; the Association of Asian Societies of Bioinformatics; Bioinformatics Institute New Zealand; and eIMBL (electronic International Molecular Biology Laboratory), an initiative of the Korean government.

Prof. Kevin Burrage and Prof. Mark Ragan are involved in the new European Union Sixth Framework project QosCosGRID (Quasi-opportunistic Supercomputing for Complex Systems on the Grid). Our European partners are ALTIA International Informatics (Hungary), Cranfield School of Management (UK), Eötvös Lorand University (Hungary), Institut national de Recherche en Informatique et en Automatique (France), Platform Computing SARL (France), Poznan Supercomputing and Networking Center (Poland), Technion – Israel Institute of Technology (Israel), Universitat Pompeu Fabra (Spain), Universiteit van Amsterdam (The Netherlands) and The University of Ulster (UK).

In March 2006, Prof. Mark Ragan made a presentation to a 14-member delegation visiting from Germany under the auspices of the Forum for European – Australian Science & Technology Co-operation (FEAST).

International visitors (2006)

- Dr Arif Anwar, Vice-President, Synamatrix Sdn Bhd, Kuala Lumpur, Malaysia
- Dr Leonard N. Blocksberg, Cartesian Gridspeed Ltd, New Zealand
- Dr Christopher Bouton, Head of Integrative Data Mining, Research Technology Center, Pfizer Inc., USA
- Dr George Chan, Sarawak State Minister of Modernisation of Agriculture, Malaysia
- Prof. John Dennis, Noah Harding Professor Emeritus, Department of Computational & Applied Mathematics, Rice University, USA
- Prof. Kim-Anh Do, Department of Biostatistics and Applied Mathematics, MD Anderson Cancer Center, USA

- Dr James J. Cai, Department of Biological Sciences, Stanford University, USA
- Dr Hidde De Jong, Directeur de recherche, INRIA, France
- Prof. Rod Downey, Department of Mathematics, Victoria University, New Zealand
- Dr Alexei Drummond, Department of Computer Science, University of Auckland, New Zealand
- Dr Victor Farutin, Research Technology Center, Pfizer Inc., USA
- Dr Paul Flicek, ENSEMBL Developer, European Bioinformatics Institute, Cambridge UK
- German delegation under auspices of FEAST (see text above)
- Prof. Michael Griebel, Institute for Applied Mathematics, University of Bonn, Germany
- Dr Murray Grigor, Director, Special Projects, University of Auckland, New Zealand
- Mr Danny Hemelin, Department of Computer Science, University of Haifa, Israel
- Prof. Michael Hendy FRSNZ, Co-Director, Allan Wilson Centre for Molecular Ecology and Evolution, Institute of Fundamental Sciences, Massey University, New Zealand
- Dr Barbara Holland, Institute of Fundamental Sciences, Massey University, New Zealand
- Dr Paul Horton, National Institute of Advanced Industrial Science and Technology, Japan
- Prof. Sang Soo Kim, Vice President (Institutes), Korea Advanced Institute of Science and Technology (KAIST), Korea
- Professor Sun Chang Kim, Director of the Institute for the BioCentury, KAIST, Korea
- Prof. Sir David Lane, Director, Institute of Molecular and Cell Biology, Singapore
- Prof. Michael Langston, Department of Computer Science, University of Tennessee, USA
- Prof. Sang Yup Lee, LG Chem Chair Professor and Director, Bioinformatics Research Centre, KAIST, Korea
- Prof. Per Lötstedt, Department of Information Technology, University of Uppsala, Sweden
- A/Prof. Catherine McCartin, Institute of Information Sciences and Technology, Massey University, New Zealand
- Dr Simon Malham, School of Mathematics, The University of Edinburgh, UK
- Mr Dan Nicolau Jr, Mathematical Institute, University of Oxford, UK
- Dr Robert Nutter, Applied Biosystems Inc., USA
- Prof. Chris Ponting, University of Oxford, UK
- Dr Andrew Roger, Canadian Institute for Advanced Research, and Department of Biochemistry and Molecular Biology, Dalhousie University, Canada
- Dr Charles Semple, Department of Mathematics and Statistics, University of Canterbury, New Zealand
- Prof. Peter Rossmanith, Department of Computer Science, University Aachen, Germany
- Prof. Zuowei Shen, Department of Mathematics, National University of Singapore
- Dr Harukazu Suzuki, RIKEN Genome Sciences Center, Japan
- A/Prof. Tin Wee Tan, Bioinformatics Centre, National University of Singapore, Singapore
- Mr Andy Watson, Senior Director, Market Development and Strategic Alliances, Applied Biosystems Inc., USA

National visitors to one or more ACB nodes (2006)

- Prof. David Abramson, Faculty of Information Technology, Monash University
- Dr Jonathan Arthur, Head, Bioinformatics Australia, Faculty of Medicine, University of Sydney
- Dr Ashley Buckle, Department of Biochemistry and Molecular Biology, Monash University
- Prof. Matthew Bellgard, Director, Centre for Comparative Genomics, Murdoch University
- Dr Michael Charleston, School of Information Technologies, University of Sydney
- Dr David Hansen, e-Health Research Centre, CSIRO
- Dr Seokhee Hong, NICTA and School of Information Technologies, University of Sydney
- Dr Mark Morrison, CSIRO Livestock Industry
- Prof. John O'Callaghan, CEO, Australian Partnership in Advanced Computing
- Dr Mick Reid, Facilitator, NCRIS Strategic Area 5.1

- Prof. Terry Speed, Department of Statistics, University of California Berkeley, USA and Walter and Eliza Hall Institutes of Medical Research, Melbourne
- Mr Bruce Stevens, Partnerships in Science Education, Education Queensland
- Mr Paul Taylor, Senior Applications Engineer, The MathWorks Australia
- Prof. Ah Chung Tsoi, Head, Monash e-Research Centre, Monash University
- Prof. Ian Turner, School of Mathematical Sciences, Queensland University of Technology
- Prof. Robert Williamson, Chief Researcher, NICTA
- Dr Andreas Zankl, Royal Childrens' Hospital, Brisbane

Sabbatical visitors (2006)

- Prof. Mark Borodovsky, Schools of Biology and Biomedical Engineering, Georgia Institute of Technology
- Dr Regina Berretta, Newcastle Bioinformatics Initiative, The University of Newcastle (to UQ)
- A/Prof. Pablo Moscato, Director, Newcastle Bioinformatics Initiative, The University of Newcastle (to UQ)
- Mr Hong Liang Hiew, Centre for Comparative Genomics, Murdoch University
- Distinguished Prof. David Lambert FRSNZ, Allan Wilson Centre for Molecular Evolution, and Institute of Molecular Biosciences, Massey University, Auckland

National and international workshops

In 2006, ACB hosted and/or co-sponsored:

- ACB 2006 Retreat (2 February 2006) – ACB hosted a 1-day retreat on 2 February 2006. All ACB Chief Investigators attended, and a limited number of external colleagues were invited. The purpose of the retreat was to prepare for the mid-term review by ARC and to consider the strategic direction for ACB in the context of the opportunity to propose that our Centre be extended through December 2010
- Winter School in Mathematical and Computational Biology (26–30 June 2006) (refer to page 48)

- ACB-BINZ Bioinformatics Student Symposium, Auckland, NZ (11–15 July 2006) (refer to page 49)
- Workshop on Intelligent Systems for Bioinformatics 2006 (WISB 2006), Hobart (4 December 2006)
- BioInfoSummer 2006, Canberra (4–8 December 2006)

Other research linkages

Our investigators and senior members of their research groups maintain active research linkages in Australia with essentially every research-active university, many other research institutions, several Divisions within CSIRO, and with companies. In addition we have active research linkages with numerous national and international institutions.

National collaborating institutions:

- Australian Centre for Plant Functional Genomics
- Australian Partnership in Advanced Computing
- CSIRO Livestock Industry
- e-Health Research Centre, CSIRO
- Griffith University
- James Cook University
- Macquarie University
- Monash University
- Queensland Department of Primary Industries and Fisheries
- Queensland Institute of Medical Research (QIMR)
- Queensland University of Technology
- The Walter and Eliza Hall Institute of Medical Research (WEHI)
- University of Sydney

International collaborating institutions:

- Capital Normal University, China
- Carleton University, Canada
- Caesaria Rothschild Institute, Israel
- Chang Gung Memorial Hospital, Taiwan
- Chinese Academy of Science, China
- Chinese University of Hong Kong, China
- City University of Hong Kong, China
- Dalhousie University, Canada
- Durham University, UK
- ETH, Switzerland
- Free University of Berlin, Germany
- Fudan University, China
- Gene Networks International Inc., Japan and UK
- Howard Hughes Research Institute Janelia Farm, USA

- Huazhong University of Science and Technology, China
- Humboldt University of Berlin, Germany
- Institute for the Mathematical Sciences, India
- Kings College, University of London, UK
- Korea Advanced Institute for Science and Technology (KAIST)
- Kyoto University, Japan
- Lebanese American University, Lebanon
- Leeds University, UK
- Leiden University Medical Centre, The Netherlands
- Massey University, New Zealand
- Max Planck Institute of Biochemistry, Germany
- McGill University, Canada
- MIT, Boston, USA
- National Institute of Child Health & Human Development (NICHD), USA
- National University of Singapore
- Netanya Academic College, Israel
- Oak Ridge National Laboratory, USA
- Ohio University, USA
- Oxford University, UK
- Oxford e-Science Centre, UK
- Renmin University, China
- Rice University, USA
- RIKEN Genome Sciences Centre, Japan
- San Diego Supercomputer Center, USA
- Salk Institute, USA
- Tamkang University, Taiwan
- Technion – Israel Institute of Technology, Israel
- Technical University of Denmark
- Technical University of Eindhoven, The Netherlands
- Tel Aviv University, Israel
- Texas A & M University, USA
- Scripps Research Institute, USA
- SAP Laboratories, China
- Universidad de Malaga, Spain
- Universidade Federal do Rio de Janeiro, Brazil
- Universitat Politècnica de Catalunya, Spain
- Université de Nantes, France
- Université Paris-Sud, France
- University College Dublin, Ireland
- University of Aachen, Germany
- University of Auckland, New Zealand
- University of Bergen, Norway
- University of Bonn, Germany
- University of British Columbia, Canada
- University of California Berkeley, USA
- University of California San Diego, USA
- University of California Santa Cruz, USA
- University of Canterbury, New Zealand
- University of Edinburgh, UK
- University of Freiburg, Germany
- University of Georgia, USA
- University of Haifa, Israel
- University of Hamburg, Germany
- University of Hawaii, USA
- University of Houston, USA
- University of Jena, Germany
- University of Karlsruhe, Germany
- University of Leipzig, Germany
- University of Maryland, USA
- University of Modena, Italy
- University of New Brunswick, Canada
- University of Newcastle-Upon-Tyne, UK
- University of Sheffield, UK
- University of Strathclyde, UK
- University of Tennessee, USA
- University of Toronto, Canada
- University of Trier, Germany
- University of Tokyo, Japan
- University of Victoria, Canada
- University of Urbino, Italy
- University of Washington, USA
- University of Waterloo, Canada
- US Department of Energy Joint Genome Institute
- US National Center for Supercomputing Applications
- Utrecht University, The Netherlands
- Valladolid University, France
- Victoria University of Wellington, New Zealand
- Waikato University, New Zealand
- Weizmann Institute of Science, Israel
- Wuhan University, China

4. End-user links

Commercialisation activities

ACB is directly involved with other commercial partners in the following ways in 2006:

- Research collaboration with IBM (PI Isidore Rigoutsos) has led to new applications for pattern discovery (including IBM's TEIRESIAS algorithm) developed in part during ACB PhD student Michael Höhl's internship at IBM TJ Watson Research Center (see our 2004 Annual Report, page 45).
- In December 2006, IMBcom and Pfizer Inc. executed a contract in support of the research collaboration between Pfizer and ACB researchers in the application of Semantic Web technologies to integration of protein-protein interaction data (Prof. Jane Hunter and Prof. Mark Ragan) as well as the modelling of signalling networks in mammalian cells (Prof. Kevin Burrage).
- Again working through IMBcom, in 2006 ACB upgraded its license of the Cell Illustrator™ software from Genetic Networks Inc. (Tokyo) to cover the newly released version 3.0. Our access includes confidential documentation, and unique links with the GNI software development team at FQS Poland (a Fujitsu partner). This level of access will allow us to integrate the Cell Illustrator™ visualisation and modelling environment into our Visible Cell™ software at the Java class library level.
- IMBcom on behalf of our Centre signed an MTA with CSIRO Mathematical & Information Sciences for the transfer of unique image data generated in our Program 1.
- ACB was actively involved in the formation of Queensland Facility for Advanced Bioinformatics (see page 40). QFAB is recruiting senior managerial personnel (CEO, Technical Manager) and has instituted a business model that builds economic sustainability over the next few years.

- QCIF (Queensland Cyber Infrastructure Foundation Ltd, previously known as Queensland Parallel Supercomputing Foundation Ltd), SGI Inc., and The MathWorks Australia sponsored our workshops and the 2006 Winter School in Mathematical and Computational Biology.
- During 2005-2006 ACB, together with partners in the tertiary education, hospital and private sectors, played an active role in the Queensland ICT R&D strategic review process. An investment proposal was approved in principle by Cabinet in May 2006, based on which a submission was made in November 2006 into the 2007 State budget process.
- Applied Biosystems (Foster City, California) announced a strategic research collaboration with The University of Queensland in next-generation DNA sequencing, particularly in support of Dr Sean Grimmond's research on transcriptional complexity of mammalian cells.
- Two companies, AITIA International Informatics Inc. (Budapest) and Platform Technologies SARL (Paris), are partners in the QosCosGRID Project funded by the European Union Sixth Framework Information Society Technologies program (see page 50).

Government, industry and business briefings

ACB has reached out to government agencies, industry and business *via* service on committees and boards, information sessions and other interactions. Examples include our proactive role in the Queensland ICT R&D strategy process, which involves universities, institutes, hospitals, government agencies and major companies; and hosting both of the NCRIS 5.1 national bioinformatics platform meetings, which involved 18 institutions and companies.

Bioinformatics Australia is established within AusBiotech, the peak body for the Australian biotechnology industry. Dr Rohan Teasdale was instrumental in founding Bioinformatics Australia. ACB alumna Dr Lucia Santoso is Executive Officer of the Australian Bioinformatics Network operated by Bioinformatics Australia.

A/Prof. Phoebe Chen

- Steering Committee, Bioinformatics Australia

Prof. Bernard Pailthorpe

- DEST, Canberra (July, August, September)
- Australian Institute of Marine Science, Townsville (March)
- Department of Education & Training, Victoria Government, Melbourne (March)
- Department of State Development, Queensland Government, Brisbane (August)
- Information Industries Bureau, Queensland Government, Brisbane (August)
- Department of Natural Resources, Queensland Government, Brisbane (August)
- Prime Minister's Science, Education and Innovation Council (PMSEIC) Working Group on Scientific Data, Canberra (December)

Prof. Mark Ragan

- Queensland Government ICT R&D Review, Brisbane (February)
- 2006 e-Health Research Colloquium Program, Brisbane (March)
- AIIA network breakfast for the ICT industry in Queensland (June)
- Ms Michele Robinson, Director, International Collaborations, Department of the Premier & Cabinet, Queensland Government (June)
- Department of State Development, Queensland Government, Brisbane (June)
- Dr Stephen Duckett, Queensland Health, Brisbane (July)
- NCRIS 5.1 National Reference Group (May, June, August, September)
- NCRIS 5.16 Data Management Workshop, Melbourne (October)
- NCRIS 5.1 Bioinformatics Capability Consultative Group (BCCG) meeting, Sydney (November)

Dr Rohan Teasdale

- Steering Committee, Bioinformatics Australia



Training in technology transfer and commercialisation

Five students from ACB's nodes took part in the IMBcom BioBusiness training course in 2006 (see page 42).

Public awareness programs

The ACB research website (<http://bioinformatics.org.au>; see also page 47) was inaugurated in 2004. Our web site:

- maintains an active, up-to-date list of Australian and overseas bioinformatics events, conferences and job opportunities
- provides links to the National Bioinformatics Strategy document, the DITR report on bioinformatics issues and opportunities for Australia, and other key national resources
- hosts a growing number of unique software and data resources in bioinformatics and computational biology. All major software tools developed in ACB are hosted (*e.g.* MEME, MAST, AltExon, LOCATE, PhosphoregDB, VariantDB, EEEP, EMU, GANN, WOOF, EvoSimulator)
- provides hyperlinks to software from other Australian bioinformatics research groups at ANU, Monash, UNSW, UQ, and the Walter & Eliza Hall Institute.
- received more than 10,000 unique hits during 2006 from a monthly average of 555 unique addresses, with slightly more than half of the visits originating from the .au (Australia) top-level domain
- served more than 49 GB of data in 2006.

Prof. Mike Fellows was awarded a development grant by Google for a project in popularisation of computer science called *Unplugged++*. Prof. Fellows is senior author of the book *Computer Science Unplugged* (see <http://www.unplugged.canterbury.ac.nz/>), which provides teachers with a series of off-line activities designed to let people of all ages have fun exploring interesting ideas in computer science. *Computer Science Unplugged* is being translated into Chinese, Greek, Korean, Norwegian and Swedish.

Working through the Partnerships in Science Education program of Education Queensland, several secondary-school science teachers enrolled in our 2006 Winter School. Other means by which we might help promote science and technology in general, and the bioscience-mathematics interface in particular, were discussed with our Advisory Board in November 2006. Several possibilities will be pursued once the 2007 secondary-school term is underway.

5. Organisational support

New organisations recruited to or involved in the Centre

During the preparation of our case for extension of ACB, we were approached by researchers and/or senior representatives of four Australian universities to discuss possible inclusion in an expanded Centre. Based on the specific programs proposed and potential levels of funding, only one could be included. At the time this Annual Report was prepared (February 2006), we are awaiting results of this competitive application round. As discussed elsewhere (pages 50 and 53), research collaborations were initiated in 2006 with Applied Biosystems and with Pfizer.

Infrastructure

World-class expertise:

Bioinformatics itself is a key research technology. ACB researchers, students and collaborators have access to top Australian experts in many areas of bioinformatics (as well as in statistics, high-performance computing, and data integration). This expertise is being aggregated further through QFAB, NCRIS and other initiatives.

Computing:

The ARC Centre in Bioinformatics is headquartered on Level 6-West of the A\$105 million Queensland Bioscience Precinct (QBP) on the St Lucia campus of UQ. QBP-6W has been purpose-built for bioinformatics, with space for more than 75 researchers and students. All ACB researchers access some of the best computing facilities in Australia, including a 16-processor IBM p690 Regatta and a 120-processor IBM Opteron cluster (IMB), and two 64- and one 16- processor SGI Altix systems (UQ), a new \$12,000 cluster at Newcastle Bioinformatics Initiative, 100s-TB-scale storage, and (*via* merit allocation) the 2100-processor APAC national computing facility. A new 60-processor cluster, purchased with support of ARC LIEF, is being installed to support a mirror of the UCSC Genome Browser.

Data infrastructure:

ACB has built significant data environments, database mirrors and specialised software, some of which are exposed on our research website. QFAB will further develop this infrastructure with commercial bioinformatics software (*e.g.* SRS) and data sources, and will be an active member of the developing national computational and data grids within the evolving national collaborative research infrastructure. We have access to storage at ANU and UQ currently at the 100s TB scale, and being developed toward PB scale; 40 TB is currently reserved for us at UQ. Both ACB and QFAB are participants in the APAC national computing and data grids.

Microscopy:

The \$1.4M ACRF/IMB Dynamic Imaging Facility and the High-content Screening facility were established with active involvement of CI Rohan Teasdale. It contains two state-of-the-art Zeiss Meta 510 confocal microscope systems including two-photon and motorised stage capabilities. This live cell imaging facility complements the Advanced Cryo-Electron Microscopy Laboratory at IMB, a major node of the Nanostructural Analysis Network Organisation MNRF that provides instruments for high-resolution 3D reconstruction and analysis of cells and macromolecules. Adjunct Research Committee member Dr Brad Marsh played a key advisory role in the acquisition of its high-end instrumentation, which remains unique in Australasia and is comparable to world-best. This laboratory is based around a 300-keV Tecnai G2 FEG transmission EM equipped with dual Gatan Ultrascan 4Kx4K CCD cameras, an embedded energy filter, compustage and high-tilt tomography holders for both ambient-temperature and frozen specimens.

Microarrays and other wet-lab technologies:

Queensland Bioscience Precinct has state-of-the-art facilities and infrastructure to undertake a wide range of experimental phenomic research through the ARC Special Research Centre for Functional and Applied Genomics, including microarray production and analysis, protein expression and crystallisation, mass spectrometry and proteomics, single-particle analysis, transgenic animals, and the zebrafish model system. Next-generation high-throughput DNA sequencing is being installed in collaboration with Applied Biosystems Inc.



Annual cash contributions: collaborating organisations

	Year 1	Year 2	Year 3	Year 4	Year 5
ARC	250,000	250,000	250,000	250,000	250,000
Relinquished ARC DP grants *	531,079	531,079	531,079	531,079	531,079
The University of Queensland **	550,000	550,000	550,000	500,000	500,000
The Australian National University	10,000	10,000	10,000	10,000	10,000
Deakin University	10,000	10,000	10,000	10,000	10,000
The University of Newcastle	10,000	10,000	10,000	10,000	10,000
IBM Australia	0***	0	0	0	0
TOTAL	1,361,079	1,361,079	1,361,079	1,311,079	1,311,079

* Possible adjustments during Years 2-5 not shown.

** In addition to the cash support identified here, the Centre will benefit from half the IGS generated by the core ARC grant funds at UQ that is distributed *via* IMB.

*** Does not include an IBM SUR (Shared University Research) grant valued at A\$3 million.

Annual in-kind contributions: collaborating organisations

	Year 1	Year 2*	Year 3*	Year 4*	Year 5*
The University of Queensland					
Personnel	490,364	507,261	523,970	539,691	555,880
Equipment	40,000	40,000	40,000	40,000	40,000
Other	55,680	0	0	0	0
Total	586,044	547,261	563,970	579,691	595,880
The Australian National University					
Personnel	33,000	33,990	35,010	36,060	37,142
Equipment	0	0	0	0	0
Other	0	0	0	0	0
Total	33,000	33,990	35,010	36,060	37,142
Deakin University					
Personnel	26,956	27,765	28,598	29,456	30,339
Equipment	0	0	0	0	0
Other	0	0	0	0	0
Total	26,956	27,765	28,598	29,456	30,339
The University of Newcastle					
Personnel	39,477	40,661	41,881	43,138	44,432
Equipment	0	0	0	0	0
Other	0	0	0	0	0
Total	39,477	40,661	41,881	43,138	44,432
IBM Australia					
Personnel	265,000	0	0	0	0
Equipment	0	0	0	0	0
Other	266,000	266,000	266,000	0	0
Total	531,000	266,000	266,000	0	0
TOTAL	1,216,477	915,677	935,459	688,345	707,793

* The numbers shown

TOTAL (Annual Cash + Annual in-kind contributions)	2,577,556	2,276,756	2,296,538	1,999,424	2,018,872
---	------------------	------------------	------------------	------------------	------------------

6. National benefit

The ARC Centre in Bioinformatics is the only research centre in Australia that offers both a central focus in bioinformatics, and presence in multiple states and institutions. With 82 staff and 14 research groups, we can effectively deliver breadth of expertise and critical mass into important problems in genome-phenome biology. Our leadership role in national and international research, infrastructure, and advanced training initiatives is extensively documented elsewhere in this Report. With the emergence of Bioinformatics Australia and the National Bioinformatics Network, and the organisation of bioinformatics capacity through NCRIS – each an initiative in which ACB has played an active part – we reaffirm our commitment to work with all stakeholders to ensure the development of bioinformatics and its effective application to the bioscience domain in Australia.

The potential of ACB research to create opportunities for industry is illustrated by the QFAB initiative (which offers secure environments and commercial software), and our collaborations with Applied Biosystems Inc., IBM, and Pfizer. Our flagship Visible Cell™ project is seen as an example of innovative research linking ICT and bioscience, and itself is attracting commercial interest in Australia. In these ways as in others, ACB research supports national Research Priority 3: Frontier Technologies for Building and Transforming Australian Industries, specifically Priority Goals *Breakthrough Science*, *Frontier Technologies*, *Smart Information Use* in the biotechnology domain.

Mr Shev MacNamara

PhD student
ARC Centre in Bioinformatics
The University of Queensland



Together with my supervisors Prof. Kevin Burrage and Dr Roger Sidje, I have been researching algorithms for the numerical solution of the chemical master equation (CME). The CME provides a discrete and stochastic modelling framework for biochemical kinetics. This perspective has become very important because it is now recognized that intrinsic noise is a critical driver of gene regulatory networks, for example manifesting itself as the stochastic switching of the lysis/lysogeny pathway in development cycle of lambda phage. Previously these equations were felt to be intractable because of the inherently high-dimensional nature of the problem. However we have recently made considerable progress *via* Krylov methods and operator splitting, and truly scalable methods for large-scale problems arising in computational biology are being developed. In particular, our work on the mitogen-activated protein kinase cascade promises a much more sophisticated understanding of cell-signaling mechanisms in a discrete and stochastic context, especially the phenomenon of ultrasensitivity that they exhibit.

I co-presented the outcomes of this work in August 2006 at the International conference on Positive Systems: Theory and Applications in Grenoble, France. This was the first time I had attended a conference further afield than Australia and New Zealand. I was then invited to present this work in a seminar at the Centre for Mathematical Biology, Oxford University, under the directorship of Prof. Philip Maini, and also at the Bioinformatics Group at University College London, on the invitation of Dr Karen Page. Academically these were wonderful opportunities for which I am grateful, while at the same time, on a personal level, being immersed in the ancient cultures and traditions of Oxford and more generally of Europe has given me a broader perspective.



Names of ACB investigators, postdoctoral researchers and students are shown in boldface. We also list a few publications that were developed with ACB support (e.g. data, computing) but do not show ACB co-authorship; these have no author names in boldface, and have not been counted toward our KPIs. Similarly we list some publications that are available on-line ahead of print, but do not count them toward our KPIs.

A. BOOKS

A3. Edited

Jiang, T., Yang, U.-C., **Chen, Y.P.P.** & Wong, L., Eds (2006) Proceedings of the 4th Asia-Pacific Bioinformatics Conference. Imperial College Press, London. Advances in Bioinformatics and Computational Biology Series, Volume 3. ISBN: 1-86094-623-2.

Rosamond, F. & Copes, L., Eds (2006) Educational transformations. Changing our lives through mathematics. A tribute to Stephen Ira Brown. Institute for Studies in Educational Mathematics, and Retreat for the Arts and Sciences, Bloomington IN, xvii + 557 pp. ISBN: 1-4208-0690-4.

B. BOOK CHAPTERS

Bordes, N., Ulm, S., Pettersen, Ø; Murphy, K., Gwynne, D., Pagnon, W., Hungerford, S., Hiscock, P., Hall J. & **Pailthorpe, B.** (2006) Data Grid for the management, reconstruction, analysis and visualisation of archaeological data. *In: An Archaeological Life: Papers in Honor of Jay Hall* (S. Ulm & I. Lilley, Eds), University of Queensland Press, pp. 251-264. Research Report Series 7. ISBN: 1086499-863-6.

Burrage, K., Burrage, P., **Hamilton, N.** & **Tian, T.** (2006) Compute intensive simulations for cellular modelling. *In: Parallel Computing for Bioinformatics and Computational Biology* (A.Y. Zomaya, Ed.), John Wiley and Sons, Chapter 4, pp. 79-120. ISBN: 978-0-471-71848-2.

Fellows, M. (2006) A short meditation on Steve Brown's main pedagogical idea, as related by Frances Rosamond. *In: Rosamond, F.* and Copes, L. (2006) Educational transformations. Changing our lives through mathematics. A tribute to Stephen Ira Brown. Institute for Studies in Educational Mathematics, and Retreat for the Arts and Sciences, Bloomington IN. pp. 400-406. ISBN: 1-4208-0690-4.

Hegland, M. (2006) An approximate maximum *a posteriori* method with Gaussian process priors. *In: Contributions to probability and statistics: applications and challenges. Proceedings of the International Statistics Workshop, University of Canberra, 4-5 April 2005*, pp. 261-275 (P. Brown, S. Liu & D. Sharma, Eds). World Scientific. ISBN: 981-270-391-8.

McLachlan, G.J. (2006) Discriminant analysis. *In: The Encyclopedia of Measurement and Statistics* (N.J. Salkind, Ed.), Thousand Oaks, California: Sage. ISBN: 978-1412916110.

Rosamond, F. (2006) A Ph.D. candidate's experience with Steve Brown's treasured videotapes. *In: Rosamond, F.* & Copes, L. (2006) Educational transformations. Changing our lives through mathematics. A tribute to Stephen Ira Brown. Institute for Studies in Educational Mathematics, and Retreat for the Arts and Sciences, Bloomington IN. pp. 235. ISBN: 1-4208-0690-4.

Rosamond, F. (2006) Off-line and on-line computer games and mathematical sciences popularization. *In: Rosamond, F.* & Copes, L. (2006) Educational transformations. Changing our lives through mathematics. A tribute to Stephen Ira Brown. Institute for Studies in Educational Mathematics, and Retreat for the Arts and Sciences, Bloomington IN. pp. 407-426. ISBN: 1-4208-0690-4.

C. JOURNAL ARTICLES

C1. Journal articles in scholarly refereed publications

Aturaliya, R.N., **Fink, J.L.**, **Davis, M.J.**, **Teasdale, M.S.**, **Hanson, K.A.**, Miranda, K.C., **Forrest, A.R.**, **Grimmond, S.M.**, Suzuki, H., Kanamori, M., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y. & **Teasdale, R.D.** (2006) Subcellular localization of mammalian Type II membrane proteins. *Traffic* 7:613-625.

Barrio M., **Burrage, K.**, **Leier A.** & **Tian T.** (2006) Oscillatory regulation of Hes1: discrete stochastic delay modelling and simulation. *PLoS Computational Biology* 2: e117.

Bailey, T.L., Williams, N., Misleh, C. & Li, W.W. (2006) MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research* 34:W369-W373.

Beiko, R.G., **Harlow, T.J.** & **Ragan, M.A.** (2006) Searching for convergence in phylogenetic Markov chain Monte Carlo. *Systematic Biology* 55:553-565.

Beiko, R.G. & **Hamilton, N.** (2006) Phylogenetic identification of lateral genetic transfer events. *BMC Evolutionary Biology* 6:15.

Bennetts, J.S., Fowles, L.F., Butterfield, N.C., Berkman, J.L., **Teasdale, R.D.**, Simpson, F. & Wicking, C. (2006) Identification and analysis of novel genes expressed in the mouse embryonic facial primordia. *Frontiers in Bioscience* 11:2631-2646.

Bodén, M. & **Bailey, T.L.** (2006) Identifying sequence regions undergoing conformational change *via* predicted continuum secondary structure. *Bioinformatics* 22:1809-1814.

Bodén, M., **Yuan, Z.** & **Bailey, T.L.** (2006) Prediction of protein continuum secondary structure with probabilistic models based on NMR solved structures. *BMC Bioinformatics* 7:68.

Bucciarelli B., **Hanan, J.**, Palmquist D. & Vance C. (2006) A standardized method for analysis of *Medicago truncatula* phenotypic development. *Plant Physiology* 142:207-219.

- Burrage, K., Hood, L. & Ragan, M.A.** (2006) Advanced computing for systems biology. *Briefings in Bioinformatics* 7:390-398.
- Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., **Frith, M.C., Forrest, A.R.**, Alkema, W.B., Tan, S.L., Plessy, C., Kodzius, R., Ravasi, T., Kasukawa, T., Fukuda, S., Kanamori-Katayama, M., Kitazume, Y., Kawaji, H., Kai, C., Nakamura, M., Konno, H., Nakano, K., Mottagui-Tabar, S., Arner, P., Chesi, A., Gustincich, S., Persichetti, F., Suzuki, H., **Grimmond, S.M.**, Wells, C.A., Orlando, V., Wahlestedt, C., Liu, E.T., Harbers, M., Kawai, J., Bajic, V.B., Hume, D.A. & Hayashizaki, Y. (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genetics* 38:626-635.
- Caruana, G., Cullen-McEwen, L., Nelson, A.L., Kostoulas, X., Woods, K., Gardiner, B., **Davis, M.J.**, Taylor, D.F., **Teasdale, R.D., Grimmond, S.M.**, Little, M.H. & Bertram, J.F. (2006) Spatial gene expression in the T-stage mouse metanephros. *Gene Expression Patterns* 6:807-825.
- Chan, C.X., Beiko, R.G. & Ragan, M.A.** (2006) Detecting recombination in evolving nucleotide sequences. *BMC Bioinformatics* 7:412.
- Chan, C.X.**, Ho, C.L. & Phang, S.M. (2006). Trends in seaweed research. *Trends in Plant Science* 11:165-166.
- Chen, Q. & **Chen, Y.P.P.** (2006) Mining frequent patterns for AMP-activated protein kinase regulation on skeletal muscle. *BMC Bioinformatics* 7:394.
- Chen, Y.P.P., Prompramote, S. & Maire, F.** (2006) MDSM: Microarray Database Schema Matching using the Hungarian method. *Information Sciences* 176:2771-2790.
- Davis, M.J., Hanson, K.A., Clark, F., Fink, J.L., Zhang, F.**, Kasukawa, T., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y. & **Teasdale, R.D.** (2006) Differential use of signal peptides and membrane domains is a common occurrence in the protein output of transcriptional units. *PLoS Genetics* 2:e46.
- Davis, M.J., Zhang, F., Yuan, Z. & Teasdale, R.D.** (2006) MemO: A consensus approach for the annotation of membrane organization. *In Silico Biology* 6:387-399.
- Dawson, P.A., Gardiner, B., **Grimmond, S.** & Markovich, D. (2006) Transcriptional profile reveals altered hepatic lipid and cholesterol metabolism in hyposulfatemic NaS1 null mice. *Physiological Genomics*. 26:116-124.
- Deutsch, E.W., Ball, C.A., Bova, G.S., Brazma, A., Bumgarner, R.E., Campbell, D., Causton, H.C., Christiansen, J., Davidson, D., Eichner, L.J., Goo, Y.A., **Grimmond, S.**, Henrich, T., Johnson, M.H., Korb, M., Mills, J.C., Oudes, A., Parkinson, H.E., Pascal, L.E., Quackenbush, J., Ramialison, M., Ringwald, M., Sansone, S.A., Sherlock, G., Stoeckert, Jr., C.J., Swedlow, J., Taylor, R.C., Walashek, L., Zhou, Y., Liu, A.Y. & True, L.D. (2006) Development of the Minimum Information Specification for In Situ Hybridization and Immunohistochemistry Experiments (MISFISHIE). *OMICS* 10:205-208.
- Fellows, M.**, Gramm, J. & Niedermeier, R. (2006) On the parameterized intractability of motif search problems. *Combinatorica* 26:149-167.
- Fellows, M.**, Dujmovic, V., Hallett, M., Kitching, M., Liotta, G., McCartin, C., Nishimura, N., Ragde, P., **Rosamond, F.**, Suderman, M., Whitesides, S. & Wood, D.R. (2006) A fixed-parameter approach to two-Layer planarization. *Algorithmica* 45:159-182.
- Fellows, M.**, Szeider, S. & Wrightson, G. (2006) On finding short resolution refutations and small unsatisfiable subsets. *Theoretical Computer Science* 351:351-359.
- Fink, J.L., Aturaliya, R.N., Davis, M.J., Zhang, F., Hanson, K.A., Teasdale, M.**, Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y. & **Teasdale, R.D.** (2006) LOCATE: a mouse protein subcellular localization database. *Nucleic Acids Research* 34(Database issue):D213-217.
- Forrest, A.R.**, Taylor, D.F., Crowe, M.L., Chalk, A.M., Waddell, N.J., Kolle, G., **Faulkner, G.J.**, Kodzius, R., Katayama, S., Wells, C., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y. & **Grimmond, S.M.** (2006) Genome-wide review of transcriptional complexity in mouse protein kinases and phosphatases. *Genome Biology* 7:R5.
- Forrest, A.R.**, Taylor, D.F., **Fink, J.L.**, Gongora, M.M., Flegg, C., **Teasdale, R.D.**, Suzuki, H., Kanamori, M., Kai, C., Hayashizaki, Y. & **Grimmond, S.M.** (2006) PhosphoregDB: the tissue and sub-cellular distribution of mammalian protein kinases and phosphatases. *BMC Bioinformatics* 7:82.
- Frith, M.C.**, Ponjavic, J., Fredman, D., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y. & Sandelin, A. (2006) Evolutionary turnover of mammalian transcription start sites. *Genome Research* 16:713-722.
- Frith, M.C., Forrest, A.R.**, Nourbakhsh, E., Pang, K.C., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., **Bailey, T.L. & Grimmond, S.M.** (2006) The abundance of short proteins in the mammalian proteome. *PLoS Genetics* 2:e52.
- Frith, M.C.**, Wilming, L.G., **Forrest, A.R.**, Kawaji, H., Tan, S.L. Wahlestedt, C., Bajic, V.B., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., **Bailey, T.L. & Huminiacki, L.** (2006) Pseudo-messenger RNA: phantoms of the transcriptome. *PLoS Genetics* 2:e23.
- Frith, M.C., Bailey, T.L.**, Kasukawa, T., Mignone, F., Kummerfeld, S.K., Madera, M., Sunkara, S., Furuno, M., Quackenbush, J., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., Pesole, G. & Mattick, J.S. (2006) Discrimination of non-protein-coding transcripts from protein-coding mRNA. *RNA Biology* 3:40-48.
- Furuno, M., Pang, K.C., Ninomiya, N., Fukuda, S., **Frith, M.C.**, Bult, C., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., Mattick, J.S. & Suzuki, H. (2006) Clusters of internally primed transcripts reveal novel long noncoding RNAs. *PLoS Genetics* 2:e37.
- Garcia, A.**, Rocca-Serra, P., Stevens, R., Taylor, C., Nashar, K., **Ragan, M.A. & Sansone, S.A.** (2006) The use of concept maps during knowledge elicitation in ontology development processes. The nutrigenomics use case. *BMC Bioinformatics* 7:267.
- Garcke, J., **Hegland, M.** & Nielsen, O. (2006) Parallelisation of sparse grids for large scale data analysis. *ANZIAM Journal* 48:11-22.
- Hamilton, N.** & Thas, J.A. (2006) Maximal arcs in PG(2,q) and partial flocks of the quadratic cone. *Advances in Geometry* 6:39-51.

- Hodge, D., Coghill, E., Keys, J., Maguire, T., Hartmann, B., McDowall, A., Weiss, M., **Grimmond, S.** & Perkins, A. (2006) A global role for EKLF in definitive and primitive erythropoiesis. *Blood* 107:3359-3370.
- Höhl, M., Rigoutsos, I. & Ragan, M.A.** (2006) Pattern-based phylogenetic distance estimation and tree reconstruction. *Evolutionary Bioinformatics Online* 2:357-373. <http://www.la-press.com/evolbio.htm>
- Huang, B., **Yuan, Z.**, Heron, B.A., Gray, B.R., Eglezos, S., Bates, J.R. & Savill, S. (2006) Distribution of 19 major virulence genes in *Legionella pneumophila* serogroup 1 isolates from patients and water in Queensland, Australia. *Journal of Medical Microbiology* 55:993-997.
- Jensen, K., Styczynski, M., **Rigoutsos, I.** & Stephanopoulos, G. (2006) A generic motif discovery algorithm for sequential data. *Bioinformatics* 22:21-28.
- Ben-Tovim Jones, L., Bean, R.W., McLachlan, G.J. & Zhu, J.X.** (2006) Mixture models for detecting differentially expressed genes in microarrays. *International Journal of Neural Systems* 16:353-362.
- Kawaji, H., **Frith, M.C.**, Katayama, S., Sandelin, A., Kai, C., Kawai, J., Carninci, P. & Hayashizaki, Y. (2006) Dynamic usage of transcription start sites within core promoters. *Genome Biology* 7:R118.
- Kerr, M., Lindsay, M., Luetterforst, R., **Hamilton, N.**, Simpson, F., Parton, R., Gleeson, P.A. & **Teasdale, R.D.** (2006) Visualisation of macropinosome maturation by the recruitment of sorting nexins. *Journal of Cell Science* 119:3967-3980.
- Lau, R.Y.K., Tang, M., Wong, O., Milliner, S.W. & **Chen, Y.P.P.** (2006) An evolutionary learning approach for adaptive negotiation agents. *International Journal of Intelligent Systems* 21:41-72.
- Lee, A.H., Wang, K., Scott, J.A., Yau, K.K.W. & **McLachlan, G.J.** (2006) Multilevel zero-inflated Poisson regression modelling of correlated count data with excess zeros. *Statistical Methods in Medical Research* 15:47-61.
- Lee, D. & Muhlberger, R.** (2006) Mobile marketing as technology intervention: a locales framework analysis. *International Journal of Mobile Marketing* 1:59-65.
- Lin, X., Liu, Q., Lu, H., Yuan, Y. & **Zhou, X.** (2006) Summarizing level-two topological relations in large spatial datasets. *ACM Transactions on Database Systems* 31:584-630.
- Lin, X., Xu, J., Zhang, Q., Lu, H., Yu, J., **Zhou, X.** & Yuan, Y. (2006) Approximate processing of massive continuous quantile queries over high speed data streams. *IEEE Transactions on Data and Knowledge Engineering* 18:5:683-698.
- Loose, C., Jensen, K., **Rigoutsos, I.** & Stephanopoulos, G. (2006) A linguistic model for the rational design of antimicrobial peptides. *Nature* 443:867-869.
- Lopez, C. & **Chen, Y.P.P.** (2006) Using object and trajectory analysis to facilitate indexing and retrieval of video. *Knowledge-Based Systems Journal* 19:639-646.
- Maeda, N., Kasukawa, T., Oyama, R., Gough, J., Frith, M., Engstrom, P.G., Lenhard, B., **Aturaliya, R.N.**, Batalov, S., Beisel, K.W., Bult, C.J., Fletcher, C.F., **Forrest, A.R.**, Furuno, M., Hill, D., Itoh, M., Kanamori-Katayama, M., Katayama, S., Katoh, M., Kawashima, T., Quackenbush, J., Ravasi, T., Ring, B.Z., Shibata, K., Sugiura, K., Takenaka, Y., **Teasdale, R.D.**, Wells, C.A., Zhu, Y., Kai, C., Kawai, J., Hume, D.A., Carninci, P. & Hayashizaki, Y. (2006) Transcript annotation in FANTOM3: mouse gene catalog based on physical cDNAs. *PLoS Genetics*. 2:e62.
- Manson D., **Hanan, J.**, Hunt M., Bristow M., Erskine P., Lamb D. & Schmidt S. (2006) Modelling predicts positive and negative interactions between three Australian tropical tree species in monoculture and binary mixture. *Forest Ecology and Management* 233:315-323.
- Martin, H.G., Ivanova, N., Kunin, V., Warnecke, F., Barry, K., McHardy, A.C., Yeates, C., He, S., Salamov, A., Szeto, E., Dalin, E., Putnam, N., Shapiro, H.J., Pangilinan, J.L., **Rigoutsos, I.**, Kyrpides, N.C., Blackall, L.L., McMahon, K.D. & Hugenholtz, P. (2006) Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nature Biotechnology* 24:1263-1269.
- Martinez, G., Georgas, K., Challen, G.A., Rumballe, B., **Davis, M.J.**, Taylor, D., **Teasdale, R.D.**, **Grimmond, S.M.** & Little, M.H. (2006) Definition and spatial annotation of the dynamic secretome during early kidney development. *Developmental Dynamics* 235:1709-1719.
- Mau, B., Glasner, J.D., **Darling, A.E.** & Perna, N.T. (2006) Genome-wide detection and analysis of homologous recombination among sequenced strains of *Escherichia coli*. *Genome Biology* 7:R44.
- McHardy, A. C., Martin, H.G., Tsigiris, A., Hugenholtz, P. & **Rigoutsos, I.** (2006) Accurate phylogenetic classification of DNA fragments based on sequence composition. *Nature Methods* 4:63-72.
- McLachlan, G.J., Bean, R.W. & Ben-Tovim Jones, L.** (2006) A simple implementation of a normal mixture approach to differential gene expression in multiclass microarrays. *Bioinformatics* 22:1608-1615.
- McLachlan, G.J.**, Ng, S.K. & **Bean, R.** (2006) Robust cluster analysis via mixture models. *Austrian Journal of Statistics* 35:157-174.
- McMillan, D.J., **Beiko, R.G.**, Geffers, R., Buer, J., Schouls, L.M., Vlamincx, B.J.M., Wannet, W.J.B., Sriprakash, K.S. & Chhatwal, G.S. (2006) Genes for the majority of group A streptococcal virulence factors and extracellular surface proteins do not confer an increased propensity to cause invasive disease. *Clinical Infectious Diseases* 43:884-888.
- Miranda, K., Huynh, T., Tay, Y., Ang, Y.-S., Tam, W.-L., Thomson, A., Lim, B. & **Rigoutsos, I.** (2006) A pattern-based method for the identification of micro-RNA target sites and their corresponding heteroduplexes. *Cell* 126, 1203-1217.
- Mitchell, E.K., Taylor, D.F., Woods, K., **Davis, M.J.**, Nelson, A.L., **Teasdale, R.D.**, **Grimmond, S.M.**, Little, M.H., Bertram, J.F. & Caruana, G. (2006) Differential gene expression in the developing mouse ureter. *Gene Expression Patterns* 6: 519-538.
- Moscato, P., Mendes, A. & Berretta, R. (2006) Benchmarking a memetic algorithm for ordering microarray data. *BioSystems* 88:56-75.

- Ng, S.K., **McLachlan, G.J.** & Lee, A.H. (2006) An incremental EM-based learning approach for on-line prediction of hospital resource utilization. *Artificial Intelligence in Medicine* 36:257-267.
- Ng, S.K., **McLachlan, G.J.**, Wang, K., **Ben-Tovim Jones, L.** & Ng, S.W. (2006) A mixture model with random-effects components for clustering correlated gene-expression profiles. *Bioinformatics* 22:1745-1752.
- Nicolau, D.V., Jr, **Burrage, K.**, Parton, R.G. & Hancock, J.F. (2006) Identifying optimal lipid raft characteristics required to promote nanoscale protein-protein interactions on the plasma membrane. *Molecular and Cellular Biology* 26:313-323.
- Pang, K.C., **Frith, M.C.** & Mattick, J.S. (2006) Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends in Genetics* 22:1-5.
- Prieto, E.** & Sloper, C. (2006) Looking at the stars. *Theoretical Computer Science* 351:437-445.
- Ragan, M.A., Harlow, T.J. & Beiko, R.G.** (2006) Do different surrogate methods detect lateral genetic transfer events of different relative ages? *Trends in Microbiology* 14:4-8.
- Ravasi, T., Suzuki, H., Pang, K.C., Katayama, S., Furuno, M., Okunishi, R., Fukuda, S., Ru, K., **Frith, M.C.**, Gongora, M.M., **Grimmond, S.M.**, Hume, D.A., Hayashizaki, Y. & Mattick, J.S. (2006) Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Research* 16:11-19.
- Rigoutsos, I.**, Huynh, T., Miranda, K., Tsiganos, A., McHardy, A. & Platt, D. (2006) Short blocks from the noncoding parts of the human genome have instances within nearly all known genes and relate to biological processes. *Proceedings of the National Academy of Sciences USA* 103:6605-6610.
- Shen, H.T., Tan, K.-L., **Zhou, X.** & Cui, B. (2006) ICICLE: A semantic-based retrieval system for WWW images. *ACM Multimedia Systems Journal* 11:438-454.
- Shen, H.T., **Zhou, X.** & Cui, B. (2006) Indexing and integrating multiple features for WWW Images. *Journal of WWW* 9:343-364.
- Shen, H.T., **Zhou, X.** & Zhou, A. (2006) An adaptive and dynamic dimensionality reduction method for high-dimensional indexing. *The VLDB Journal*, doi:10.1007/s00778-005-0167-3 (Online First™, regular paper in 2007:16:2:219-234.)
- Shi, W., Zhou, W. & **Chen, Y.P.P.** (2006) Statistical analysis of TATA box and its extensions in the promoters of human genes. *Current Bioinformatics* 1:337-345.
- Song, J. & **Burrage, K.** (2006) Predicting residue-wise contact orders in proteins by support vector regression. *BMC Bioinformatics* 7:425.
- Song, J., **Burrage, K., Yuan, Z.** & Huber, T. (2006) Prediction of cis/trans isomerization in proteins using PSI-BLAST profiles and secondary structure information. *BMC Bioinformatics* 7:124.
- Song, J., Wang, M. & **Burrage, K.** (2006) Exploring synonymous codon usage preferences of disulfide-bonded and non-disulfide bonded cysteines in the *E. coli* genome. *Journal of Theoretical Biology* 241:390-401.
- Sprenger, J., Fink, J.L. & Teasdale, R.D.** (2006) Evaluation and comparison of mammalian subcellular localization prediction methods. *BMC Bioinformatics* 7:5:S3.
- Stanley, S.M., **Bailey, T.L.** & Mattick, J.S. (2006) GONOME: measuring correlations between GO terms and genomic positions. *BMC Bioinformatics* 7:94.
- Tian, T. & Burrage, K.** (2006) Stochastic models for regulatory networks of the genetic toggle switch. *Proceedings of the National Academy of Sciences of the United States of America* 103:22:8372-8377.
- Tian, T., Xu, S., Gao, J. & Burrage, K.** (2006) Stimulated maximum likelihood method for estimating kinetic rates in gene expression. *Bioinformatics* doi:10.1093/bioinformatics/btl552 (Online First™, regular paper in 2007:23:84-91.)
- Tian, T., Burrage, P.M., Burrage, K. & Carletti, M.** (2006) Stochastic delay differential equations for genetic regulatory networks. Special Issue of *Journal of Computational and Applied Mathematics* doi:10.1016/j.cam. 2006.02.063.
- Waddell, N., Jonnalagadda, J., Marsh, A., Grist, S., Jenkins, M., Hobson, K., Taylor, M., Lindeman, G.J., Tavtigian, S.V., Suthers, G., Goldgar, D., Oefner, P.J.; kConFab Investigators; Taylor, D., **Grimmond, S.**, Khanna, K.K. & Chenevix-Trench, G. (2006) Characterization of the breast cancer associated ATM 7271T>G (V2424G) mutation by gene expression profiling. *Genes Chromosomes Cancer* 45:1169-1181.
- Wells, C.A., Chalk, A.M., **Forrest, A.**, Taylor, D., Waddell, N., Schroder, K., Himes, S.R., Faulkner, G., Lo, S., Kasukawa, T., Kawaji, H., Kai, C., Kawai, J., Katayama, S., Carninci, P., Hayashizaki, Y., Hume, D.A. & **Grimmond, S.M.** (2006) Alternate transcription of the Toll-like receptor signalling cascade. *Genome Biology* 7:R10.
- Wu, H.J., Seib, K.L., Srikhanta, Y.N., Kidd, S.P., Edwards, J.L., Maguire, T.L., **Grimmond, S.M.**, Apicella, M.A., McEwan, A.G. & Jennings, M.P. (2006) PerR controls Mn-dependent resistance to oxidative stress in *Neisseria gonorrhoeae*. *Molecular Microbiology* 60:401-416.
- Xiang, L., Lee, A.H., Yau, K.K.W. & **McLachlan, G.J.** (2006) A score test for zero-inflation in correlated count data. *Statistics in Medicine* 25:1660-1670.
- Xu, K., **Zhou, X.**, Shen, H.T., Lin, X. & Deng, K. (2006) A multiresolution terrain model for efficient visualization query processing. *IEEE Transactions on Data and Knowledge Engineering* 18:1382-1396.
- Yuan, Z., Zhang, F., Davis, M.J., Bodén, M. & Teasdale, R.D.** (2006) Predicting the solvent accessibility of transmembrane residues from protein sequence. *Journal of Proteome Research* 5:1063-1070.
- Zhu, X., Ambrose, C. & McLachlan, G.J.** (2006) Selection bias in working with the top genes in supervised classification of tissue samples. *Statistical Methodology* 3:29-41.

C2. Journal article – other contribution to refereed journal

Chen, Y.P.P. & Wong, L. (2006) Algorithms and computational bioinformatics. *Journal of Bioinformatics and Computational Biology* 4:1:v-vii.

Chen, Y.P.P. & Chua, T.-S. (2006) Special issue on multimedia modelling. *The Visual Computer* 22(5):1–2.

C3. Journal articles – non refereed articles

Chen, Y.P.P., Hwang, M.J., Lee, S.K. & Wong, L. (Eds) (2006) Special issue "Bioinformatics in Asia Pacific". *Asia Pacific Biotech News* 10:1385-1503.

Fellows, M. (2006) Parameterized complexity and stability of approximation: a tale of two programs. *Parameterized Complexity Newsletter* (F. Rosamond, Ed.) 2(11):4-7.

Fellows, M., Christian, R., **Rosamond, F.** & Slinko, A.M. (2006) On the complexity of lobbying in multiple referenda. *Cahiers de recherche du The Centre for Interuniversity Research in Qualitative Economics (CIREQ), Montreal, Canada* 13-2006:1-7.

Ragan, M.A. (2006) ARC Centre in Bioinformatics. *Asia Pacific Biotech News* 10:1391-1394.

Teasdale, R.D. & **Hamilton, N.** (2006) Imaging the future: managing the exponential growth in cell image data. *Australian Biochemist* 37:16-20.

E1. Conference – full written paper – refereed proceedings

An, J. & **Chen, Y.P.P.** (2006) Finding rule groups to classify high dimensional gene expression datasets. *In: Proceedings of the 18th International Conference on Pattern Recognition 2006, Hong Kong, 20-24 August 2006*, pp.1196-1199 (Y.Y. Tang, S.P. Wang, G. Lorette, D.S. Yeung & H. Yan, Eds). IEEE CS Press. ISBN: 0-7695-2521-0.

An, J. & **Chen, Y.P.P.** (2006) Finding short patterns to classify text documents. *In: Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI-06), Hong Kong, 18-22 December 2006*, pp. 293-296 (T. Nishida, Z. Shi, U. Visser, X. Wu, J. Liu, B. Wab, W. Cheung & Y.M. Cheung, Eds). IEEE CS Press. ISBN: 0-7695-2747-7.

An, J. & **Chen, Y.P.P.** (2006) A similarity search algorithm to predict protein structures. *In: Proceedings of 10th International Conference on Knowledge-Based & Intelligent Information & Engineering Systems (KES2006) UK, 9-11 October 2006. Lecture Notes in Computer Science* 4252:1305-1312 (B. Gabrys, R.J. Howletter & L.C. Jain, Eds). Springer-Verlag. ISBN: 978-3-540-46537-9.

An, J. & **Chen, Y.P.P.** (2006) Depth-first rule generation for text categorization. *In: Proceedings of the Fourth International Conference on Active Media Technology (AMT 2006), Brisbane, Australia, 7-9 June 2006*, pp. 302-306 (Y. Li, M. Looi & N. Zhong, Eds). ISO Press. ISBN: 1-58603-615-7.

Asadi, S., Xu, J., Shi, Y., Diederich, J. & **Zhou, X.** (2006) Calculation of target locations for web resources. *In: Proceedings of the 7th International Conference on Web Information Systems Engineering (WISE 2006) Wuhan, China, 23-26 October 2006. Lecture Notes in Computer Science* 4255:277-288 (A. Aberer, Z. Peng, E.A. Rundensteiner, Y. Zhang, X. Li, Eds). Springer. ISBN: 3-540-48105-2.

Basford, K.E., **McLachlan, G.J.** & **Bean, R.W.** (2006) Issues of robustness and high dimensionality in cluster analysis. *In: Proceedings of COMPSTAT 2006, Rome, Italy, 28 August-2 September 2006* (A. Rizzi & M. Vichi, Eds), pp. 3-15. Springer. ISBN: 978-3-7908-1708-9.

Bodlaender, H.L., **Fellows, M.**, Langston, M.A., **Ragan, M.A.** & Rosamond F.A. (2006) Kernelization for convex recoloring of trees. *In: Proceedings of Algorithms and Complexity in Durham (ACiD), Durham UK, 18-20 September 2006. Texts in Algorithmics* 7:23-36.

Burrage, K., **Fellows, M.**, Estivill-Castro, V., Langston, M., **MacNamara, S.** & **Rosamond, F.** (2006) The Undirected Feedback Vertex Set Problem has polynomial kernel size. *In: Proceedings of the International Workshop on Parameterised and Exact Complexity 2006, Zürich, 13–15 September 2006. Lecture Notes in Computer Science* 4169:192-202. (H.L. Bodlaender, M.A. Langston, Eds). Springer-Verlag. ISBN: 978-3-540-39098-5.

Burrage, K., **Hegland, M.**, **MacNamara, S.** & Sidje R.B. (2006) A Krylov-based finite state projection algorithm for solving the chemical master equation arising in the discrete modelling of biological systems. *In: Proceedings of the Markov 150th anniversary conference, Charleston, SC, USA, 12-14 June 2006* (A.N. Langville & W.J. Stewart, Eds), pp.21-38. Bosen Books. ISBN: 1-932482-35-0.

Chen, Y.P.P. & Chen, Q. (2006) Analyzing inconsistency toward enhancing integration of biological molecular databases. *In: Proceedings of the Fourth Asia-Pacific Bioinformatics Conference (APBC2006), Taipei, 13-16 February 2006*, pp.197-206 (T. Jiang, U.-C. Yang, Y.P.P. Chen & L. Wong, Eds). Imperial College Press, London. ISBN: 978-1-86094-623-3.

Chen, Q., **Chen, Y.P.P.**, Zhang C. & Zhang S. (2006) Detecting collusion attacks in security protocols. *In: Proceedings of the Eighth Asia Pacific Web Conference, Harbin, China, 16-18 January 2006. Lecture Notes in Computer Science* 3841:297-306 (X. Zhou, J. Li, H.T. Shen, M. Kitsuregawa & Y. Zhang, Eds). Springer-Verlag. ISBN: 3-540-31142-4.

Chen, Q., **Chen, Y.P.P.**, Zhang, C. & Li, L. (2006) Mining frequent itemsets for protein kinase regulation. *In: Proceedings of the 9th Pacific Rim International Conference on Artificial Intelligence, Guilin, China, 7-11 August 2006. Lecture Notes in Computer Science* 4099:222-230 (Q. Yang & G. Webb, Eds). Springer-Verlag. ISBN: 978-3-540-36667-6.

Christian, R., Fellows, M., **Rosamond, F.** & Slinko, A.M. (2006) On the complexity of lobbying in multiple referenda. *In: Proceedings of the First International Workshop on Computational Social Choice (COMSOC 2006), Amsterdam, 6-8 December 2006* (U. Endriss & J. Lang, Eds) pp. 87-98.

Deng, K., **Zhou, X.**, Shen, H.T., Xu, K. & Lin, X. (2006) Surface k-NN Query Processing. *In: Proceedings of 22nd International Conference on Data Engineering (ICDE 2006), Atlanta, Georgia, USA, 3-7 April 2006*. (L. Liu, A. Reuter, K.-Y. Whang & J. Zhang, Eds), p. 78. IEEE Computer Society Press. ISBN: 0-8186-8289-2.

- Fellows, M.,** Dehne, F., Fernua, H., **Prieto, E. & Rosamond, F.** (2006) Nonblocker: parameterized algorithms for minimum dominating set. *In: Proceedings of the Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM 2006), Merin, Czech Republic, 21-27 January 2006. Lecture Notes in Computer Science* 3831:237-245 (J. Wiedermann, G. Tel, J. Pokorný, M. Bieliková & J. Stuller, Eds). Springer. ISBN: 3-540-31198-X.
- Fellows, M.** (2006) The lost continent of polynomial time: preprocessing and kernelization. *In: Proceedings of the International Workshop on Parameterised and Exact Computation 2006, Zürich, 13–15 September 2006. Lecture Notes in Computer Science* 4169:276-277. (H.L. Bodlaender & M.A. Langston, Eds). Springer-Verlag. ISBN: 978-3-540-39098-5.
- Fellows, M.,** Downey, R. & McCartin, C. (2006) Parameterized approximation problems. *In: Proceedings of the International Workshop on Parameterised and Exact Computation 2006, Zürich, 13–15 September 2006. Lecture Notes in Computer Science* 4169:121-129. (H.L. Bodlaender & M.A. Langston, Eds). Springer-Verlag. ISBN: 978-3-540-39098-5.
- Fellows, M., Rosamond, F.,** Rotics U. & Szeider, S. (2006) Cliquewidth minimization is NP-hard. *ACM Symposium on Theory of Computing (STOC 2006), Seattle, 21-23 May 2006, 354-362* (J.M. Kleinberg, Ed.). ISBN: 1-59593-134-1.
- Garcia, V., França, P., Mendes, A. & Moscato, P. (2006) A parallel memetic algorithm applied to the total tardiness machine scheduling problem. *In: Proceedings of 20th IEEE International Parallel and Distributed Processing Symposium (IPDPS 2006), Rhodes Island, Greece, 25–29 April, 2006. IEEE Computer Society Press. ISBN: 1-4244-0054-6.*
- Huang, Z. & Zhou, X.** (2006) Protein structure matching by patch signatures. *In: Proceedings of the 17th International Conference and Workshop on Database and Expert Systems Applications (DEXA 2006), Krakow, 4-8 September 2006. Lecture Notes in Computer Science* 4080:528-537 (S. Bressan, J. Köng & R. Wagner, Eds). Springer. ISBN: 3-540-37871-5.
- Huang, H., Zhou, X.,** Song, D. & Bruza, P. (2006) Dimensionality reduction in motif-signature based protein structure matching. *In: Proceedings of the 17th Australasian Database Conference (ADC 2006), Tasmania, 16-19 January 2006* (G. Dobbie & J. Bailey, Eds), pp. 89-97. ISBN: 1-920682-31-7.
- Hamilton, N., Pantelic, R., Hanson, K., Fink, J.L., Karunaratne, S. & Teasdale, R.D.** (2006) Automated sub-cellular phenotype classification: an introduction and recent results. *In: Proceedings of the 2006 Workshop on Intelligent Systems for Bioinformatics (WISB 2006), Hobart, 73:67-72* (M. Bodén & T.L. Bailey, Eds). Australian Computer Society. ISBN: 1-920-68254-6.
- Kam, S.C., Lee, S.D., Park, H.C. & **Zhou, X.** (2006) Web-based system for electronic navigational charts using GML and SVG. *In: Proceedings of the 5th IEEE/ACIS International Conference on Computer and Information Science, and 1st IEEE/ACIS International Workshop on Component-Based Software Engineering, Software Architecture and Reuse (ICIS-COMSTAR'06), Hawaii, 12 July 2006.* (N. Ishii & R. Lee, Eds), pp. 485-491. IEEE Computer Society Press. ISBN: 0-7695-2613-6.
- Leier, A.,** Kuo, P.D., Banzhaf, W. & **Burrage, K.** (2006) Evolving noisy oscillatory dynamics in genetic regulatory networks. *In: Proceedings of the 9th European Conference on Genetic Programming, EuroGP 2006* (P. Collet, M. Tomassini, M. Ebner & S. Gustafson, Eds). *Lecture Notes in Computer Science* 3905:290-299. Springer. ISBN: 3-540-33143-3.
- Mahata, P., Costa, W.,** Cotta, C. & Moscato, P. (2006) Hierarchical clustering, languages and cancer. *In: Proceedings of EvoWorkshops 2006, Budapest, 10-12 April 2006. Lecture Notes in Computer Science* 3907:67-78 (F. Rothlauf, J. Branke, S. Cagnoni, E. Costa, C. Cotta, R. Drechsler, E. Lutton, P. Machado, J.H. Moore, J. Romero, G.D. Smith, G. Squillero & H. Takagi, Eds). Springer. ISBN: 3-540-33237-5.
- Ng, S.K., **McLachlan, G.J., Bean, R.W.** & Ng, S.W. (2006) Clustering replicated microarray data via mixtures of random effects models for various covariance structures. *In: Proceedings of 2006 Workshop on Intelligent Systems for Bioinformatics (WISB2006), Hobart, 73:29-33* (M. Bodén & **T.L. Bailey**, Eds). Australian Computer Society. ISBN: 1-920-68254-6.
- Ng, S.K., Wang, K. & **McLachlan, G.J.** (2006) Multilevel modelling for inference of genetic regulatory networks. *In: Proceedings of SPIE 2005, Complex Systems in the International Symposium on Microelectronics, MEMS, and Nanotechnology, Brisbane, 11-15 December 2005.* (A. Bender, Ed.) 6039:1-12. Bellingham, Washington: International Society for Optical Engineering. ISBN: 9780819460837.
- Shu, Y., Zhang, J.F. & **Zhou, X.** (2006) A Grid-enabled architecture for geospatial data sharing. *In: Proceedings of the 2006 IEEE Asia-Pacific Conference on Services Computing (APSCC 2006) Guangzhou, China, 12-15 December 2006* (Z. Chen, C.A. Shoniregun & Y.-C. You, Eds), pp. 369-375. IEEE Computer Society Press. ISBN: 0-7695-2751-5.
- Subramaniam, C., **An, J. & Chen, Y.P.P.** (2006) Microarray analysis for gene expression that involves trace metals. *In: Proceedings of the Fourth International Conference on Active Media Technology (AMT 2006), Brisbane, 7-9 June 2006,* pp. 211-216 (Y. Li, M. Looi & N. Zhong, Eds). ISO Press. ISBN: 1-58603-615-7.
- Xu, G., Zhang, Y. & **Zhou, X.** (2006) Discovering task-oriented usage pattern for Web recommendation. *In: Proceedings of the 17th Australasian Database Conference (ADC 2006), Tasmania, 16-19 January 2006* (G. Dobbie & J. Bailey, Eds), pp. 167-174. ISBN: 1-920682-31-7.

E2. Conference – full written paper – non-refereed proceedings

Garcia Castro, A., Norena, A., Betancourt, A. & **Ragan, M.A.** (2006) Cognitive support for an argumentative structure during the ontology development process. *In: Proceedings of Protégé-9*, Stanford CA, 23-26 July 2006, pp. 85-88.

Inostroza-Ponta, M., Berretta, R., Mendes, A. & Moscato, P. (2006) An automatic graph layout procedure to visualize correlated data. *In: Proceedings of the IFIP AI 2006 - IFIP Conference on Artificial Intelligence*, Santiago, Chile, 21-24 August 2006 (D. Kumar & A. Turner, Eds), pp. 179-188. Springer. ISBN: 0-387-34627-9.

E4i. Conference – edited volume of conference proceedings

Barga, R. & **Zhou, X.**, Eds (2006) Proceedings: 22nd International Conference on Data Engineering Workshops (ICDE'06), Atlanta, 3-7 April 2006. *IEEE Computer Society*. ISBN: 0-7695-2570-9

Zhou, X., Li, J., Shen, H.T., Kitsuregawa, M. & Zhang, Y., Eds (2006) Frontiers of World Wide Web Research and Development. Proceedings of APWeb 2006, Harbin, China, 16-18 January 2006. *Lecture Notes in Computer Science* 3841, Springer. ISBN: 3-540-31142-4.

E4. Conference – unpublished presentations

This listing does not include presentations at the 2005 Winter School in Mathematics and Computational Biology (see pages 48-49).

Dun, E.A., **Hanan, J.** & Beveridge, C.A. (2006) Branching regulation in pea: growing understanding with the aid of computational modelling. Combined Annual Conferences of the Australian Society for Biochemistry & Molecular Biology, the Australia and New Zealand Society for Cell and Developmental Biology and the Australian Society of Plant Scientists 2006, Brisbane, 24-28 September 2006.

Hamilton, N. (2006) Analysing real-time video microscopy: the dynamics and geometry of vesicles and tubules in endocytosis. Combined Annual Conferences of the Australian Society for Biochemistry & Molecular Biology, the Australia and New Zealand Society for Cell and Developmental Biology and the Australian Society of Plant Scientists 2006, Brisbane, 24-28 September 2006.

Hanan, J., Beveridge, C., Cox, M., Ferguson, B. & Gresshoff, P. (2006) Modelling processes involving long-distance communication in plants. Combined Annual Conferences of the Australian Society for Biochemistry & Molecular Biology, the Australia and New Zealand Society for Cell and Developmental Biology and the Australian Society of Plant Scientists 2006, Brisbane, 24-28 September 2006.

Hanan, J., Thornby, D. & Gresshoff, P. (2006) A computational approach to autoregulation of nodulation. International Conference on Legume Genetics and Genomics, Brisbane, 9-13 April 2006.

Moscato, P. (2006) Molecular signatures of breast cancer subtypes using combinatorial optimization methods. Hunter Medical Research Institute (HMRI) Conference on Translational Cancer Research, Newcastle, Australia 20-22 September 2006.

Wenden, B., Dun, E.A., **Hanan, J.**, Weller, J., Beveridge, C.A. & Rameau, C. (2006) Computational analysis of flowering in pea. Combined Annual Conferences of the Australian Society for Biochemistry & Molecular Biology, the Australia and New Zealand Society for Cell and Developmental Biology and the Australian Society of Plant Scientists 2006, Brisbane, Australia, 24-28 September 2006.

G. COMPUTER SOFTWARE

Beiko, R.G. & **Hamilton, N.** (2006) EEPEP: Efficient Evaluation of Edit Paths. <http://bioinformatics.org.au/eeep/>

Bodén, M. **Yuan, Z.** & **Bailey, T.L.** (2006) Prediction of protein continuum secondary structure with probabilistic models based on NMR solved structures. <http://www.prowler.itee.uq.edu.au/sspred>

Fink, J.L. & **Hamilton, N.A.** (2006) DomainDraw: A macromolecular schematic drawing program. <http://domaindraw.imb.uq.edu.au/>

Hamilton, N., **Pantelic, R.**, **Hanson, K.**, **Fink, J.L.**, Karunaratne S. & **Teasdale, R.D.** (2006) ASPiC: the Automated Subcellular Phenotype Classification pipeline, available on request.

Stanley, S.M., **Bailey, T.L.** & Mattick, J.S. (2006) GONOME: measuring correlations between GO terms and genomic positions. <http://gonome.imb.uq.edu.au>

OTHER ACADEMIC OUTPUT

Posters

Frith, M.C. & Bailey, T. (2006) Motif discovery with arbitrary insertions and deletions. 14th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB 2006), Fortaleza, Brazil, 6-10 August 2006.

Hamilton, N., Woodcroft, B., Burrage, K., Ragan, M.A., Torda, A. & Huber, T. (2006) Protein contact prediction using patterns of correlation. The 7th Critical Assessment of Techniques for Protein Structure Prediction, Pacific Grove, CA, 26-30 November 2006.

Mendes, A., **Mahata, P.** & Moscato, P. (2006) Semi-supervised hierarchical classification for breast cancer outcome prediction. Hunter Medical Research Institute Conference on Translational Cancer Research, Newcastle, 20-22 September 2006.

Mendes, A., Berretta, R. & Moscato, P. (2006) Benchmarking the ordering of microarray data. Hunter Medical Research Institute Conference on Translational Cancer Research, Newcastle, 20-22 September 2006.

Reports

McLaren, C.E., Gordeuk, V.R., Chen, W.-P., Harris, E.L., Snively, B.M., Adams, P.C., Barton, J.C., Acton, R.T., Reboussin, D.M., Speechley, M., Castro, O., **McLachlan, G.J., Bean, R.W.** & McLaren, G.D. (2006) The distribution of transferrin saturation and serum ferritin concentration in multiple race/ethnicities: the hemochromatosis and iron overload screening (HEIRS) study. University of California and The University of Queensland.

Ng, S.K. & **McLachlan, G.J.** (2006) Extension of mixture-of-experts networks for binary classification of hierarchical data. The University of Queensland.

Theses

Geard, N. (2006) Artificial ontogenies: a computational model of the control and evolution of development. PhD thesis, School of Information Technology and Electrical Engineering, The University of Queensland. http://www.itee.uq.edu.au/~nic/_papers/Geard06PhDThesis.pdf

Haji, P. (2006) The numerical analysis of the recovery of information from stress-strain measurements of linear viscoelastic processes. PhD thesis, Centre for Mathematics and its Applications, The Australian National University. 90 pp.

Höhl, M. (2006) Is multiple sequence alignment required for accurate inference of phylogeny? PhD thesis, Institute for Molecular Bioscience, The University of Queensland. 178 pp.

Prieto, E. (2005) Systematic kernelization in FPT algorithm design. Department of Computer Science, The University of Newcastle. 168 pp.

Wakabayashi, M., (2006). A biologically plausible model of *C. elegans* locomotion and motor control. Honours thesis, School of Information Technology and Electrical Engineering, The University of Queensland. 88 pp.

Watson, J. (2006) From genotypes to phenotypes and back again: modelling the interaction between individual behaviour and evolution. School of Information Technology and Electrical Engineering, The University of Queensland. 147 pp. (Jointly with ARC Centre in Complex Systems).

Willadsen, K. (2006) Robustness in Boolean models of genetic regulatory systems. School of Information Technology and Electrical Engineering, The University of Queensland. 256 pp. (Jointly with ARC Centre in Complex Systems).



Appendix 2: Financial statement

INCOME		2006
Total balance carry forward		757,073
ARC Centre Grant		843,452
Host Institution Support		500,000
TOTAL INCOME		<u>2,100,525</u>
EXPENDITURE		
Salaries		1,193,709
Scholarships		65,319
Student Programming/Research Experience/Workshops		36,513
Funds to other nodes		222,218
Strategic Visitors/Sponsorship Program/Director Strategic Discretionary		30,167
Equipment		23,554
Accommodation		1,646
Travel (airfares)		5,721
Maintenance/Consumables		
Annual Report	14,514	
Computer Consumables/Maintenance	662	
Laboratory Consumables	42,908	
Meetings – Advisory Board meetings & other meetings	1,622	
Photography/Slides/Posters	152	
Postage & Freight	300	
Printing – books/journals/reprints	4,396	
Relocation	2,000	
Recruitment	6,501	
Registration: trademark, domain name	4,086	
Stationery	100	
Telephone & Fax	31	
Transport – Local	2,737	
Travel Subsistence (Local and Overseas)	<u>1,782</u>	81,792
Others (hospitality, payroll tax, FBT, cost recoveries/journal adjustments etc)		<u>-34,738</u>
TOTAL EXPENDITURE		<u>1,625,902</u>
BALANCE		<u>474,623</u>







Acknowledgements:

Cover and page 32 image courtesy of Dr Brad Marsh. Originally published in the Inaugural Article: Organellar relationships in the Golgi region of the pancreatic beta cell line, HIT-T15, visualized by high resolution electron tomography. *Proceedings of the National Academy of Sciences USA* (February 27, 2001) 98:2399-2406.

Images shown on pages 30 and 36 were provided courtesy of Dr Markus Kerr and Dr Rohan Teasdale, and those on pages 27 and 34 by Adam Costin, Garry Morgan, Andrew Noske and Dr Brad Marsh, Institute for Molecular Bioscience, The University of Queensland.



Australian Government
Australian Research Council

ARC Centre in Bioinformatics
Institute for Molecular Bioscience
The University of Queensland
Brisbane Qld 4072

Telephone:
(07) 3346 2617

Facsimile:
(07) 3346 2101

Email:
acb@imb.uq.edu.au

Website:
<http://bioinformatics.org.au>

Design by:
PROdesign Australia

