Contents

CHAIR’S FOREWORD 2
DIRECTOR’S REVIEW AND HIGHLIGHTS 3
ORGANISATION AND GOVERNANCE 5
Organisational chart 5
Advisory Board 6
Research Committee 9
Director 9
Chief Investigators 9
Partner Investigators 14
Special advisor 14
Strategic partnerships 15
Information infrastructure 17
Centre Manager 17
Postdoctoral Fellows / Research Officers 17
Research Assistants 17
Masters/PhD students 17
Honours students 18
International & national interns 19
Undergraduate research trainees and volunteers 19
Secondary school research project students 19
Visiting scholars, associates 19
STRATEGIC DIRECTION (2008-2010) 21
Vision 21
Mission 21
Aims 21
Objectives 21
Capabilities — core technologies and infrastructure 21
RESEARCH PROGRAMS 22
Program 1 – Experimental phenomics 23
1.1 Computational and microarray-based analysis of transcriptional complexity of mammalian cells 23
1.2 Computational cellular biology: combining computational prediction and experimental validation 24
1.3 Generation of the experimental data for the modelling of cellular compartments and the development of organelle biogenesis phenotype screens 25
Program 2 – Modelling cellular networks 26
2.1 Hybrid algorithms for large-scale network and graph optimisation problems in bioinformatics 26
2.2 Computational discovery of gene-regulatory motifs and motif clusters 28
2.3 New statistical approaches in bioinformatics 30
2.4 Mathematical modelling of gene regulatory networks 31
2.5 Complex systems modelling of regulatory and protein-protein interaction networks 32
2.6 Ortholog mapping of prokaryotic and eukaryotic genomes 33
Program 3 – the Visible Cell™ – data integration and visualisation 35
3.1 Advanced data management and integration 35
3.1.1 Data environment for the Visible Cell™ project 35
3.1.2 Integrating database technologies and visual analysis for genome data 37
3.2 Visualisation 38
3.2.1 Concept and engineering of the Visible Cell™ 38
3.2.2 Modelling and visualisation of biological processes in complex spatial environments 40
KEY PERFORMANCE INDICATORS 42
1. Research findings and competitiveness 43
Publications 43
Quality of publications 43
Number of publications 43
Patents and trademarks 43
Invitations to address and participate in conferences 44
International conferences 44
National conferences 46
Invitations to visit leading international laboratories, major research groups and centres 46
Commentaries about the Centre’s achievements 47
Evidence of interdisciplinary linkages within the Centre 47
2. Research Training and professional education 48
Recruitment 48
Professional/technical training and advanced career development 48
Seminar/Symposium hosted and/or sponsored by ACB 48
CILR/ACB Joint Seminar 48
2007 Winter School in Mathematical and Computational Biology 48
Bioinformatics Australia 2007 Conference 49
2007 ACB–BINZ Bioinformatics Student Symposium 49
ACB 2007 Annual All-Hands Meeting 50
Other research and training and professional education 50
Postgraduate and undergraduate courses in the Centre’s area 50
Senior high school student research scheme 50
3. International, national and regional links and networks 51
Visitors 51
International visitors 51
National visitors to one or more ACB nodes 51
Sabbatical visitors 51
Collaborative national and international workshops and exchanges 52
Other research linkages 53
Membership on national and international professional committees 54
Research projects with international partners 54
4. End-user links 56
Commercialisation activities 56
Government, industry and business briefings 56
Trained/ing in technology transfer and commercialisation 57
Public awareness programs 57
5. Organisational support 58
New organisations recruited to or involved in the Centre 58
Infrastructure 58
Annual cash contributions from other organisations 59
Annual in-kind contributions: collaborating organisations 59
6. National benefit 60
Measures of expansion of Australia’s capability in the priority area(s) 60
APPENDIX 1: PUBLICATIONS 61
APPENDIX 2: FINANCIAL STATEMENT 70
It gives me considerable pleasure to introduce you to the 2007 Annual Report of the ARC Centre in Bioinformatics (ACB). This report marks the conclusion of the initial funding phase for Australia’s first multi-institutional research centre in bioinformatics. However, we are very gratified that the Australian Research Council (ARC) this past year has agreed to award funding for the continuation of ACB through to the end of 2010. Furthermore, the ARC determined to upgrade the status of the ACB to that of an ARC Centre of Excellence.

Three additional institutions have joined the renewed ACB: Macquarie University, the University of Auckland, and the University of Tennessee. This second round of support allows the Centre to build on its existing research portfolio and to reinforce the capabilities and interests of its Investigators and their research teams across both existing and new interdisciplinary research programs and between geographical nodes.

During the past year, the Centre continued to achieve outstanding outcomes. Among these were the following key achievements:

- ACB organised the 2007 Winter School in Mathematical and Computational Biology.
- It was involved in numerous ways with Bioinformatics Australia, AusBiotech and the annual conference BA2007.
- ACB set up the second joint Student Symposium in Bioinformatics with Bioinformatics Institute New Zealand (BiNZ). It sponsored the venue, local arrangements and accommodation for New Zealand participants.
- ACB also continued to grow its highly regarded research publication record, the details of which are provided later in this annual report.
- The Centre enabled the installation of two next-generation SOLiD DNA sequencing platforms in Dr Sean Grimmond’s laboratory at The University of Queensland (UQ).
- ACB also established the BioMANTA collaboration (www.biomanta.org) between Pfizer Research Technology Center (Cambridge, Massachusetts) and UQ.
- It became a partner in the QosCosGRID project that is now funded by the European Union Sixth Framework Information Society Technologies program.
- ACB has also been very active in the establishment of the Queensland Facility for Advanced Bioinformatics.

Most importantly, ACB’s research excellence continues to be recognised internationally. Examples include Professor Kevin Burrage’s acceptance of a joint appointment as Professor of Computational Systems Biology at Oxford University. Professor Mike Fellows took up a senior Humboldt Fellowship that enables him to make extended working visits to European research centres in mathematics and computational biology. Professor Geoff McLachlan accepted an appointment to the ARC College of Experts.

We anticipate that 2008 will be another year of considerable achievement for ACB as it commences activities under its second term of support from ARC.

Dr Gregory R. Smith
Chair
March 2008
I'm equally delighted to report that ARC concomitantly upgraded the status of our Centre to that of an ARC Centre of Excellence. This superb outcome reflects the splendid dedication and commitment of everyone associated with our Centre: our institutional partners, Advisory Board, Investigators, postdoctoral fellows, postgraduate students and professional staff. Special thanks are due to Ms Nicole Thompson, Executive Officer in the Office of the Deputy Vice-Chancellor (Research) at The University of Queensland, and to Ms Lanna Wong, Manager of our Centre, for professional support of the very highest level.

Four of our Investigators have chosen not continue with us into the extended Centre: Dr Sean Grimmond, who is leading the introduction into Australia of the Applied Biosystems next-generation SOLiD DNA sequencing; Dr Ralf Muhlberger, who has taken on directorship of the Information Environments programme of the School of Information Technology and Electrical Engineering at UQ; Professor Bernard Pailthorpe, who has added interim chairmanship of the Australian Research Collaboration Service (ARCS) under the National Collaborative Research Infrastructure Strategy (NCRIS) to his role as CEO of Queensland Cyber Infrastructure Foundation; and Dr Prasad Kodali, who has been promoted to America Sales Leader for Software Solutions within IBM Healthcare and Life Sciences. I also acknowledge Dr Jennifer Hallinan, one of our founding Investigators, who in November 2005 took up an appointment at the University of Newcastle-upon-Tyne in the UK. I thank each of these colleagues for their many contributions to the Centre, and am delighted that many collaborations with these colleagues will continue.

We will be joined in our Centre of Excellence by four new Chief Investigators (Dr Brad Marsh, Professor John Mattick, A/Professor Pablo Moscato, and Professor Shoba Ranganathan) and two new Partner Investigators (Professor Michael Langston and Professor Allen Rodrigo). Five of these distinguished researchers already have relationships with ACRS – indeed, their photographs have appeared in one or more of our previous Annual Reports. I welcome these colleagues into our Centre of Excellence, and refer you to their brief academic biographies at pages 15–16.

New members of our Advisory Board are Professor Peter Donnelly FRS, University of Oxford; Professor Barney Glover, DVC(R), The University of Newcastle; and Professor Brandon Wainwright, Director, Institute for Molecular Bioscience, UQ. I warmly thank former Advisory Board members Professor John Mattick, who is assuming a new role as a CI, and Professor Lap-Chee Tsui, who is stepping down owing to his heavy commitments as Vice-Chancellor of Hong Kong University, for their valuable advice to the Centre and to me personally during the initial phase of the Centre.

With the extension, our Centre has begun to implement a revised research programme. Some previous projects have been terminated or are being phased out, new ones are being introduced, and most others are being renewed. In making this transition we must keep our eye on the challenges facing researchers in a multi-genome world. Many of our previous directions have served us well, including our foci on scaleable algorithms, workflows, advanced data integration, computational modelling and cellular image data. Our extended Centre will place greater emphasis on integration, networks and systems. This is easier said than done, but it is deeply gratifying our Investigators have already decided that two of our four programs must work as one.
Research excellence is the foundation of our Centre, and during 2007 several of our Investigators received well-deserved recognition. ARC Federation Fellow Professor Kevin Burrage accepted a joint appointment as Professor of Computational Systems Biology at Oxford University and will spend several months there each year, opening unique opportunities for collaboration and postgraduate student exchange. Professor Mike Fellows took up a senior Humboldt Fellowship that will enable extended working visits to European research centres in mathematics and computational biology. Professor Geoff McIachlan accepted appointment to the ARC College of Experts. Dr Sean Grimmond and Dr Rohan Teasdale took up NHMRC Senior Research Fellowships.

During 2007 our investigators contributed to key national and international initiatives in bioinformatics, e-research and related areas. Within Australia these include NCRIS BioPlatforms and Platforms for Collaboration, the Queensland Facility for Advanced Bioinformatics, the Australian Bioinformatics Network (managed by an ACB alumna, Dr Lucia Santoso), Bioinformatics Australia (our national peak body, incorporated within AusBiotech), the ARC Research Network in Enterprise Information Infrastructure, and the ARCHER e-research project. Internationally we collaborate in AASBi (Asia-Pacific), BioMANTA (US), FANTOM (Japan), GosCoxGRID (Europe) and others. New opportunities may arise in consequence of Australia’s new associate membership in the European Molecular Biology Laboratory (EMBL).

From the beginning, ACB has been committed to advanced education and training in bioinformatics and related disciplines. During 2007 we led the organisation of the annual Bioinformatics Australia conference, hosted the second Australia-New Zealand student bioinformatics symposium, and held our fourth annual Winter School in Mathematical and Computational Biology. Graduate students from our Newcastle and UQ nodes participated in the annual BioBusiness Retreat organised by IMBcom.

2008 will bring the 19th International Conference on Genome Informatics (GIW-2008) to Australia. This is the world’s longest-running bioinformatics conference, and the 2008 meeting will be its first in Australia, indeed only its second outside Japan. An exciting program is being planned, with distinguished keynotes and open roundtables focused on leading-edge technology, innovation, and the expansion of genome-scale bioinformatics in China and India. ACB research will also be highlighted at Pattern Recognition in Bioinformatics (PRIB-08) and eResearch Australasia 2008.

As we celebrate the extension of our Centre and the successes of our colleagues within and external to ACB, we must not lose sight of the challenges that lie ahead for bioinformatics in Australia. New technologies allow the generation of data at rates scarcely imaginable even two years ago; we will soon have not four human genomes (as today) but thousands, then perhaps millions. All these data – from sequencers, expression platforms, molecular characterisation and imaging – must be captured, managed, integrated, assigned semantic value, reasoned-over, represented graphically, modelled and simulated if we are to understand cells and organisms as systems; and this is before we interface genome-phenome data with geographical information systems, agricultural breeding programs, medical imaging, hospital and public-health records, or real-time biosecurity awareness networks. We have much to offer, and much to learn from, our colleagues in other disciplines and areas of expertise.

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

Professor Mark Ragan
Director
ARC Centre of Excellence in Bioinformatics
March 2008
Organisation and Governance

Organisational Chart
Advisory Board

The 7th Advisory Board meeting was held on 27 April 2007, and a special progress meeting was convened on 27 November 2007.

There were changes to the composition of the Advisory Board:

- Prof. Lap-Chee Tsui, Vice-Chancellor, University of Hong Kong has stepped down.
- Prof. John Mattick will change his role to that of Chief Investigator from January 2008.
- Prof. Brandon Wainwright, Director, Institute for Molecular Bioscience, Prof. Barney Glover, Deputy Vice-Chancellor (Research), University of Newcastle, and Prof. Peter Donnelly, Director, Wellcome Trust Centre for Human Genetics, University of Oxford will join our Advisory Board from January 2008.

As of 31 December 2007, 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 co-founder and director of SciVentures Investments Pty Ltd, the manager of the SciVentures Pre-Seed Fund. He also is a director or chair of four 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. He is the Board Chair for the Australian Microscopy and Microanalysis Research Facility. Greg 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. Peter Donnelly FRS
   Director, Wellcome Trust Centre for Human Genetics
   Professor of Statistical Science
   University of Oxford, UK
   Professor Donnelly is Director of the Wellcome Trust Centre for Human Genetics and Professor of Statistical Science at the University of Oxford. He is a Fellow of the Royal Society and has been awarded the Mitchell Prize of the ASA and International Society for Bayesian Analysis, and the Guy Medal in Silver of the Royal Statistical Society. His early work was in applied probability, particularly stochastic models in genetics, but for some years his research has focused on genetics. He played a major role in the HapMap project, an international collaboration that followed the Human Genome Project in studying genetic diversity in worldwide populations. He currently chairs the Wellcome Trust Case Control Consortium, a collaboration of over 200 UK scientists studying the genetics of 12 common human diseases. Other research interests include human population structure and histories, bacterial variation, human recombination, and the development of statistical methodology for modern genetics.
4 Prof. Simon Easteal
Co-Director, Centre for Bioinformatics 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 how the evolutionary dynamic between humans and their environments shaped the complexity of human biology, gave rise to human diversity and left a lasting impact on human health. 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 has served on numerous committees and boards in both private and public sectors.

5 Prof. Barney Glover
Deputy Vice-Chancellor (Research)
The University of Newcastle
Newcastle, NSW, Australia

Prof. Glover holds a Bachelor of Science (Honours), a Diploma of Education, a Master of Science and a Doctor of Philosophy from the University of Melbourne, and was a Professorial Fellow with the Western Australian Centre of Excellence in Industrial Optimization. Professor Glover also has extensive industry networks and has held a number of corporate appointments including board membership of the Australian Biosecurity Cooperative Research Centre, the John Curtin Institute for Public Policy, and the Cooperative Research Centre for Coal in Sustainable Development.

6 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. 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. Mark is Vice-President of Bioinformatics Australia, and chair (2008) of the Executive group of the Association of Asian Societies of Bioinformatics. Under the overall theme of comparative and computational genomics, research in his group includes projects on whole-genome duplication, genetic recombination, lateral genetic transfer, the cellular localisation of molecular-interaction networks, and applications of semantic web technologies to large-scale biomolecular, chemical and structural data.
7 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 chaired the National Committee of DVCs/PVCs (Research) in 2001 and was convener of the Go8 DVCs(R) until December 2007. Currently, David is a Director of the Australian Synchrotron Company and Australian Synchrotron Holding Company; AHURI Queensland Research Centre Ltd; and CRCMining.

8 Prof. Brandon Wainwright  
Director  
Institute for Molecular Bioscience  
The University of Queensland  
Brisbane QLD, Australia  

Professor Wainwright was appointed Director, Institute for Molecular Bioscience (IMB) in 2006, and before that was IMB’s Deputy Director (Research). IMB, recognised nationally and internationally as one of Australia’s leading centres for molecular bioscience research, has around 500 research staff and research students with an annual budget in the order of $40m. Professor Wainwright currently serves on a number of boards, including those of the Australian Genome Research Facility, the Australian Phenomics Facility, and the Health and Medical Research Council of Queensland. In 2007 he was Chair of the DEST Research Quality Framework Panel 6.  

Professor Wainwright’s major focus in his laboratory is the use of genomic approaches to dissect the basis of common genetic disease, in particular cystic fibrosis and basal-cell carcinoma of the skin. Through the mapping and isolation of the genes which are responsible for these diseases, he has continued to follow-through on each to understand how the genetic defects lead to the disease. The ultimate aim of his work is to discover the genes which when altered cause the tumour cell to grow in an uncontrolled fashion. Ultimately this will provide validated targets against which potential therapeutics can be developed.

9 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 80 publications, and editor of two books. He is the Director of the NSW Systems Biology Initiative. Marc is a Senior Editor of the journal Proteomics, and serves on the Editorial Boards of Practical Proteomics and Bioinformatics and Biology Insights and on the Advisory Board of the Australian Proteomics Computational Facility (APCF) as well as the Management Committee of the Ramaciotti Centre for Gene Function Analysis. He has six years experience in industry, and has co-founded two biotechnology companies, one of which, Proteome Systems, is now listed on the Australian Stock Exchange and employs about 40 staff. His current research interests are in the dynamics of the interactome and the role that protein post-translational modifications play in the control of protein-protein interactions.
Research Committee
The Research Committee met 12 times during 2007. To facilitate the transition to Phase Two of the ARC Centre of Excellence in Bioinformatics, the Director invited Dr Brad Marsh, A/Prof. Pablo Moscato and Prof. Shoba Ranganathan onto the Research Committee in 2007. They join the Centre as Chief Investigators from 2008.

Chair Prof. Mark Ragan
Advisory Board Member Prof. John Mattick
Program 1 Dr Rohan Teasdale
Program 2 Prof. Kevin Burrage
Program 2 Prof. Geoff McLachlan
Program 2 A/Prof. Pablo Moscato (Adjunct from July 2007)
Program 3 A/Prof. Phoebe Chen
Program 3 Dr Brad Marsh (Adjunct to June 2007)
Program 3 Prof. Shoba Ranganathan (Adjunct from July 2007)
Secretary Ms Lanna Wong

Director
1 Prof. Mark Ragan FLS
   Institute for Molecular Bioscience
   The University of Queensland
   Please refer to page 7 for profile.

Chief Investigators
2 Dr Timothy L. Bailey
   Institute for Molecular Bioscience
   The University of Queensland
   Dr Timothy Bailey has been active in bioinformatics and computational biology since the early 1990s, and since 2004 has been a senior research fellow at the Institute for Molecular Bioscience. He is well-known as an expert in pattern recognition in biological sequences, and as the author of several important bioinformatics software programs, most notably MEME, one of the most heavily used tools in sequence analysis. Tim’s major areas of expertise are machine learning, statistical pattern recognition (discovery), and the statistics of sequence alignment scores. He is experienced in developing parallel algorithms for massively parallel computer architectures, and in development and support of web-based algorithms for bioinformatics. His MEME website was used by approximately 1000 researchers per month in 2007, compared with 800 in 2005. In the past year, Dr Bailey’s group has developed two additional sequence analysis tools, DE ME and GLAM2.
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), Director of the Advanced Computational Modelling Centre (ACMC) at UQ, and Director of the ViSAC visualisation laboratory. In 2003 he was awarded a prestigious Federation Fellowship by the Australian Research Council. Kevin has co-authored more than 170 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. In a computational biological context he is interested in the modelling and simulation of complex cellular processes. In 2007 Kevin accepted a joint appointment as Professor of Computational Systems Biology at Oxford University.

4 A/Prof. Yi-Ping Phoebe Chen
School of Engineering Information Technology
Deakin University

Associate Professor (Reader) Yi-Ping 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. A/Prof. Chen 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. She has written about 120 refereed journal articles and international conference papers including papers in Nucleic Acids Research and BMC Bioinformatics, 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 also serves 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 cryptography, algebraic graph theory, and the popularisation of mathematical sciences. A recipient in 2006 of a Humboldt Research Award, and in 2007 of an Inaugural Fellowship to the Durham University Institute of Advanced Studies, Prof. Fellows spent much of 2007 with research groups in Europe (Durham University, University of Bergen, ETH Zürich, and Friedrich Schiller University in Jena) as an invited distinguished visiting researcher.
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 CJ Martin Career Development Award. He joined the Institute for Molecular Bioscience in 2004 as head of IMB’s microarray facility. Sean was named Eppendorf Young Australian Scientist 2004, and in 2007 took up an NHMRC Senior Research Fellowship. 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 is President of the Australasian Microarray and Associated Technologies Association (AMATA). During 2006 Sean was scientific lead on a new strategic partnership with Applied Biosystems Inc. that has introduced the ABI next-generation SOLiD sequencing platform to Australia.

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  
Australian Professorial Fellow  
Professor and Head, Mathematics  
The University of Queensland  

Prof. Geoff McLachlan is internationally known for his research in statistics applied in the fields of classification, cluster and discriminant analysis, image analysis, intelligent systems, machine learning, neural networks, pattern recognition, and statistical inference. 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, focussing 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, and in 2007 was appointed to the ARC College of Experts.
9 Prof. John Mattick AO
Professor of Molecular Biology and ARC Federation Fellow
Institute for Molecular Bioscience
The University of Queensland

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. His main research interest is the role of non-protein-coding RNA in the evolution and development of humans and other complex organisms. He was awarded the 1989 Pharmacia-LKB Biotechnology Medal from 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, and elected as an Associate Member of the European Molecular Biology Organisation in 2007. John serves on the advisory boards of several institutes in Australia and abroad.

10 Dr Ralf M. Muhlberger
School of Information Technology & Electrical Engineering
The University of Queensland

Dr Ralf Muhlberger is Director of the Information Environments programme of the School of IT at UQ. He 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 Director of the Interaction Design and Multimedia programmes 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).

11 Prof. Bernard A. Pailthorpe
School of Physical Sciences
The University of Queensland; and CEO, Queensland Cyber Infrastructure Foundation

Prof. Bernard Pailthorpe is foundation Chair of Computational Science at The University of Queensland. Since 2003 has been CEO of Queensland Cyber Infrastructure Foundation (QCF), the state partner within the Australian Partnership in Advanced Computing (now the National Computing Infrastructure within NCRIS Platforms for Collaboration). In 2007 he was named interim Chair of the Australian Research Collaboration Service, also within NCRIS PIC. Before returning to Australia in 2000 to develop Sydney VisLab, Bernard was Director for Visualisation at the San Diego Supercomputer Centre at University of California San Diego, and Associate Director of the US National Science Foundation’s 50-university National Partnership in Advanced Computing Infrastructure. He is building computational, data and visualisation infrastructure and skills in support of research both nationally and within Queensland, including ACB and QFAB. Bernard’s research interests are in statistical mechanics, data and image processing, and scientific visualisation.
12 Dr Rohan D. Teasdale
Institute for Molecular Bioscience
The University of Queensland

Dr Rohan Teasdale holds a Senior Research Fellowship from the National Health & Medical Research Council. 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 played a key role in the formation of Bioinformatics Australia and is a member of its steering committee.

13 Prof. Xiaofang Zhou
School of Information Technology & Electrical Engineering
The University of Queensland

Prof. Xiaofang Zhou is Professor of Computer Science at UQ, and Head of the Data and Knowledge Engineering program in the School of IT and EE. He is also Convenor of the ARC Research Network in Enterprise Information Infrastructure (EII), and Chair of the Steering Committee of the Asia-Pacific Web Conferences. Until 1999, Xiaofang was Principal Researcher and Head of Spatial Information Systems within the Commonwealth Scientific and Industrial Research Organisation (CSIRO). Xiaofang is author of over 100 publications in spatial databases and information systems, Grid computing, and bioinformatics.

Chief Investigators Meeting on 24 April 2007. From left: Dr Rohan Teasdale, Mr Oliver Cairncross, Prof. Mark Ragan, Dr Timothy Bailey, Prof. Geoff McLachlan, A/Prof. Pablo Moscato, A/Prof. Phoebe Chen, Prof. Kevin Burrage, Prof. Xiaofang Zhou, Prof. Shoba Ranganathan, Dr Markus Hegland and Prof. John Mattick.
Partner Investigators

14 Dr Prasad Kodali
IBM Healthcare & Life Sciences
Yorktown Heights, New York USA

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.

15 Dr Isidore Rigoutsos
IBM Thomas J. Watson Research Center
Yorktown Heights, New York USA

Dr Isidore Rigoutsos manages the Bioinformatics and Pattern Discovery group at IBM’s Thomas J. Watson Research Center, 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.

16 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 Kannelakis 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 Zurich. Gene has been an Honorary Professor in the Institute for Molecular Bioscience at UQ since 2003.
Strategic partnerships

ACB is delighted that our close partnerships with the following researchers have been leveraged into the extension of our Centre’s funding from ARC. The investigators based in Australia will join our Centre of Excellence as Chief Investigators, and those overseas as Partner Investigators.

17 Prof. Michael Langston

Department of Computer Science
University of Tennessee, USA

Professor Mike Langston’s research interests include computational biology, graph theory, discrete optimization and high-performance computation. He is currently Professor of Electrical Engineering and Computer Science at the University of Tennessee, and Collaborating Scientist in the Biosciences Division at Oak Ridge National Laboratory. Mike is lead investigator on a variety of research projects focused on the design and synthesis of combinatorial algorithms used to process high-throughput biological data. His collaborations extend across the US and to research groups in Australia, Canada, Europe and the Middle East. He has worked closely for over twenty years with Professor Mike Fellows, and has authored over 200 refereed journal articles, conference papers, book chapters and other reports. Beginning in 2007, Mike led the development of the Bertinoro Systems Biology series of workshops. His work has been funded in the US by the National Science Foundation, the National Institutes of Health, the Department of Defence and the Department of Energy, and internationally by the Australian Research Council and the European Commission.

18 Dr Brad Marsh

Senior Research Fellow and Group Leader
Institute for Molecular Bioscience
The University of Queensland

19 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 was the founding director of the Newcastle Bioinformatics Initiative (2002) and is now Co-Director of the Centre for Bioinformatics, Biomarker Discovery and Information-based Medicine of The University of Newcastle, Australia. He is a member of the editorial board of Journal of Mathematical Modelling and Algorithms, Journal of Heuristics, and Memetic Computing. Pablo introduced memetic algorithms in 1989 when he was a member of the Caltech Concurrent Computation Program, and has successfully applied these hybrid computing methodologies to large-scale optimisation problems. Pablo serves on the program committee of numerous international conferences in heuristics and optimisation.

20 Prof. Shoba Ranganathan
Department of Chemistry and Biomolecular Sciences & Biotechnology Research Institute
Macquarie University

Prof. Shoba Ranganathan holds the first Chair in Bioinformatics in Australia, at Macquarie University. She holds an Adjunct Professor position at the National University of Singapore (since 2004) and a UNESCO Chair in Biodiversity Informatics. Her research interests include genome annotation, transcriptome analysis, structural bioinformatics, immunoinformatics and genome-phenome analysis. Shoba’s achievements include the development of graph theoretical methods for alternative splicing analysis, a rapid and accurate docking protocol for vaccine design, software tools for the analysis of protein sequence and for modelling protein structure, and the development of boutique databases. She pursues collaborations with research groups at the National University of Singapore, the Institute for Infocomm Research Singapore, the Johns Hopkins University, and the Dana Farber Cancer Institute at Harvard University. Shoba serves on the editorial boards of several bioinformatics journals including Briefings in Bioinformatics and BMC Bioinformatics, and contributes to the organisation and scientific program of several international bioinformatics conferences.

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

Allen Rodrigo is Professor of Computational Biology and Bioinformatics, and Director of New Zealand’s Bioinformatics Institute at the University of Auckland. He has over 80 international publications on bioinformatics and computational biology, phylogenetics and evolutionary genetics, and the molecular evolution of viruses. Allen is an Associate Editor of Evolutionary Bioinformatics, sits on the Scientific Advisory Board of two bioinformatics companies, is an Associate Investigator of the Allan Wilson Centre for Molecular Ecology and Evolution, and is involved in national and international collaborative projects on genomics and bioinformatics. His major research contributions are in the area of virus evolutionary genetics, where he has spearheaded the development of new methods to analyse time-series genetic data from viral populations.

Group Leaders on the occasion of the special Advisory Board meeting, 27 November 2007.
From left: Prof. Mark Ragan, A/Prof. Phoebe Chen, Prof. Brandon Wainwright, Prof. John Mattick, Dr Greg Smith, A/Prof. Pablo Moscato, Prof. Geoff McLachlan, Dr Markus Hegland and Prof. Kevin Burrage.
Information infrastructure
Mr Matthew Bryant
System Administrator, High-performance computing
Mr Oliver Cairncross
Project Manager, Visible Cell™
Dr Lynn Fink (to April 2007)
Research Website Administrator
Ms Mhairi Marshall
Grid & Database Developer
Mr Tim McComb
Application 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 (to March 2007)
Dr Mikael Bodén
Dr Pamela Burridge
Dr Aaron Darling
Dr Melissa Davis
Dr Lynn Fink (to April 2007)
Dr Nicholas Hamilton
Dr Jim Hanan
Dr John Hawkins
Dr Karin Kassahn
Dr Krzysztof Kurowski
Dr Fran Rosamond
Dr Osvaldo Anibal Rosso
Dr Muhammad Shoaiib Sehgal
Dr Neelima Pottekkat Sidharthan
Dr Tianhai Tian (to July 2007)
Dr James Watson
Dr Simon Wong
Dr Ian Wood
Dr Zheng Yuan

Research Assistants
Mr Cheong Xin Chan
Mr Lloyd Flack
Ms Seetha Karunarathe
Ms Chikako Ragan
Mr Benjamin Woodcroft

Masters/PhD students

Supervision/Associate supervision (principal advisor in ACB):

Farah Abdullah
(PhD thesis submitted, January 2008)
Advisor: Prof. Kevin Burrage

Rajith Aturaliya
(PhD thesis submitted, March 2008)
Advisor: Dr Rohan Teasdale

Denis Bauer (due to complete in 2009)
Advisor: Dr Timothy Bailey

Michael Bode (due to complete in 2008)
Advisor: Prof. Kevin Burrage

Durgaprasad Bollina (PhD awarded, 2008)
Advisor: Prof. Shoba Ranganathan

Alhadi Bustamam (due to complete in 2008)
Advisor: Prof. Shoba Ranganathan

Associate Advisors: Dr Nicholas Hamilton & Dr Shoaiib Sehgal

Elsa Chacko (due to complete in 2010)
Advisor: Prof. Shoba Ranganathan

Cheong Xin Chan (PhD awarded, 2008)
Advisor: Prof. Mark Ragan

Associate Advisor: Dr Robert Beiko

Feng Chen (due to complete in 2010)
Advisor: A/Prof. Phoebe Chen

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

Khar Heng Choo (due to complete in 2008)
Advisor: Prof. Shoba Ranganathan

Associate Advisor: A/Prof. Tan Tin Wee

JooYoung Choi (due to complete in 2009)
Advisor: Prof. Mark Ragan

Associate Advisor: Dr Melissa Davis

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

Wagner Emanoel Costa (due to complete in 2008)
Advisor: A/Prof. Pablo Moscato

Associate Advisor: Dr Regina Berretta

Adam Costin (due to complete in 2010)
Advisor: Dr Brad Marsh

Melissa Davis (PhD awarded, 2007)
Advisor: Dr Rohan Teasdale

Associate Advisor: Dr Zheng Yuan

Geoffrey Faulkner (due to complete in 2008)
Advisor: Dr Sean Grimmond

Alex Foo (due to complete in 2010)
Advisor: Dr Brad Marsh

Alex Garcia (PhD awarded, 2007)
Advisor: Prof. Mark Ragan

Associate Advisors: A/Prof. Phoebe Chen (to 2005) & Dr Lindsay Hood

Jitendra Gaikwad (due to complete in 2010)
Advisor: Prof. Shoba Ranganathan

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

Mo’ath Adeib Hourani (PhD awarded, 2007)
Advisor: A/Prof. Pablo Moscato

Helen Zi Huang (PhD awarded, 2007)
Advisor: A/Prof. Xiaofang Zhou

Associate Advisor: Prof. Peter Bruza
(Queensland University of Technology)

Javed M. Khan (due to complete in 2010)
Advisor: Prof. Shoba Ranganathan

Varun Khanna (due to complete in 2010)
Advisor: Prof. Shoba Ranganathan

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

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

Lessheng Kong (PhD awarded 2007)
Advisor: Prof. Shoba Ranganathan

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

Associate Advisor: Dr Regina Berretta

Dennis Lee (PhD awarded, 2007)
Advisor: Dr Ralf Muhlberger

Associate Advisors: Dr Matthew Simpson & 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 2008)
Advisor: Dr Ralf Muhlberger

Associate Advisor: Dr Stephen Viller

Shiev MacNamara (due to complete in 2008)
Advisor: Prof. Kevin Burrage

Scott Mann (due to complete in 2009)
Advisor: A/Prof. Phoebe Chen

Katrina Monico (due to complete in 2008)
Advisor: Prof. Geoff McLachlan
Ann Morrison (due to complete in 2008)  
Advisor: Dr Ralf Muhlberger  
Associate Advisor: Dr Peta Mitchell  
Shivashankar H. Nagaraj  
(due to complete in 2008)  
Advisor: Prof. Shoba Ranganathan  
Associate Advisor: Prof. Robin Gasser  
Andrew Noske (due to complete in 2008)  
Advisor: Dr Brad Marsh  
Associate Advisors: Prof. Mark Ragan & Prof. Kevin Burrage  
Tatiana Semenova (PhD awarded, 2007)  
Advisor: Dr Markus Hegland  
Chang Jin Shin (due to complete in 2009)  
Advisor: Prof. Mark Ragan  
Associate Advisors: Dr Simon Wong (to 2007) & Dr Melissa Davis  
Josefine Sprenger (due to complete in 2009)  
Advisor: Dr Rohan Teasdale  
Associate Advisor: Dr Nicholas Hamilton  
Chinmu Subramanian  
(due to complete in 2012)  
Advisor: A/Prof. Phoebe Chen  
Imran Syed (due to complete in 2009)  
Advisor: Prof. Bernard Paitthorpe  
Sharnsheer Syed  
(part-time, due to complete in 2012)  
Advisor: Prof. Xiaofang Zhou  
Joo Chuan Tong (PhD awarded 2007)  
Advisor: Prof. Shoba Ranganathan  
Peter van der Heide (due to complete in 2008)  
Advisor: Dr Brad Marsh  
Lawrence Wee (due to complete in 2008)  
Advisor: Prof. Shoba Ranganathan  
Associate Advisor: A/Prof. Tan Tin Wee  
Tom Whitington (due to complete in 2011)  
Advisor: Dr Timothy Bailey  
Jason Yang (due to complete in 2008)  
Advisor: Dr Ralf Muhlberger  
Associate Advisor: Dr Stephen Viller  
Emily Zhou (due to complete 2008)  
Advisor: Prof. Xiaofang Zhou  
Associate Advisor: Dr Heng Tao Shen  
Jia Zhu (part-time, due to complete in 2011)  
Advisor: Prof. Xiaofang Zhou  
Justin Xi Zhu (due to complete in 2008)  
Advisor: Prof. Geoff McLachlan  
Associate Advisor: Dr Angus Ng  
Co-supervision (principal advisor not an ACB investigator)  
Agnes Boskovitz (PhD thesis submitted, 2007)  
Advisor: Dr Peta Mitchell  
Shivashankar H. Nagaraj  
(due to complete in 2008)  
Advisor: Dr Markus Hegland  
Principal Advisor: Dr Rajeev Goré  
Fabian Buske (Masters awarded, 2007)  
Advisor: Dr Timothy Bailey  
Principal Advisor: Dr Mikael Bodén  
Simon Carter (due to complete in 2008)  
Advisor: Dr Paul Strooper  
Principal Advisor: Dr André Leier  
Elizabeth Dun (PhD thesis submitted, 2007)  
Advisor: Prof. Bernard Paitthorpe  
Principal Advisor: A/Prof. Stuart Phinn  
Luke Hammond (due to complete in 2009)  
Advisor: Dr Nicholas Hamilton  
Principal Advisor: Prof. Jenny Stow  
Principal Advisor: Dr Rohan Teasdale & Prof. Janet Wiles  
Mitchell Stanton-Cook  
(due to complete in 2009)  
Advisor: Dr Nicholas Hamilton  
Principal Advisor: Dr Thomas Huber  
Stuart Stephen (due to complete in 2008)  
Advisor: Prof. Mark Ragan  
Principal Advisor: Prof. John Mattick  
Honours students  
Farhan Ahammed  
Slowing Concorde  
Advisor: A/Prof. Pablo Moscato  
Andrés Esteban-Marcos  
A time machine for evolution  
Advisor: Dr Aaron E. Darling  
Andrew Noble  
Slowing Concorde  
Advisor: A/Prof. Pablo Moscato  
Mitchell Stanton-Cook  
Location and orientation of ligand molecules by paramagnetic NMR techniques for application in pharmaceutical screening  
Advisors: Dr Nicholas Hamilton & Dr Thomas Huber  

Organisation and Governance
International & national interns

International

Zuzana Cienikova
National Institute of Applied Sciences (France)
Advisor: Dr Timothy Bailey

Tim Geggert
University of Frankfurt (Germany)
Advisors: Dr Nicholas Hamilton & Dr Thomas Huber

Frijtof Heyde
Fachhochschule Weihenstephan (Germany)
Advisor: Dr Jim Hanan

Arnaab Mandal
Indian Institute of Technology, Kharagpur (India)
Advisor: Prof. Mark Ragan

Himanshu Prakash
Indian Institute of Technology, Kanpur (India)
Advisor: Prof. Xiaofang Zhou

Michael Rivoire
French Institute of Forestry - ENGREF (France)
Advisor: Dr Jim Hanan

Yuan Shi
Wuhan University (China)
Advisor: Prof. Xiaofang Zhou

Triinu Tasa
Chalmers University (Sweden)
Advisors: A/Prof. Pablo Moscato & Prof. Rodney Scott

Boxuan Zhai
Wuhan University (China)
Advisor: Prof. Xiaofang Zhou

National

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

Undergraduate research trainees and volunteers

Robert Cope
Use of Gene Ontology annotation to identify functionally similar clusters of gene products
Advisor: Dr Melissa Davis

Vinh Dang
Human copy number variation and haploinsufficient genes
Advisor: Dr Karin Kassahn

Nathan Healey
Subcellular image processing
Advisor: Dr Nicholas Hamilton

Liam McIntyre
Gene copy number variation and disease in human
Advisor: Dr Karin Kassahn

Benjamin Woodcroft
Automated object based co-localisation from whole cell imaging
Supervisor: Dr Nicholas Hamilton

Secondary school research project students

- Ann Bui, Brisbane State High School
- Rohan Datar, Brisbane State High School
- Katja Kahn-Manne, The Gap State High School
- Alex Wilson, Grace Lutheran College

Supervisors: Dr Melissa Davis & Dr Simon Wong

Visiting scholars, associates

Visiting scholars

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

Dr Meike Leuger, Bielefeld University (Germany)

Dr Zemin Pan, Shihezi University School of Medicine (China)

Mr Todd J. Treangen, Technical University of Catalonia, Barcelona (Spain)

Associates

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

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

Prof. J. Thomas August, John Hopkins University School of Medicine (USA)

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

Dr Robert Beiko, Faculty of Computer Science, Dalhousie University (Canada)

Dr Mikael Bodén, School of ITEE, The University of Queensland

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

Prof. Vladimir Brusic, Director of Bioinformatics, Dana Farber Cancer Institute, Boston (USA)

Mr Geoff Bulow, Queensland Department of Primary Industries and Fisheries

Dr Robert Charlebois, Sanofi Pasteur (Canada)

Dr Brian Dalrymple, CSIRO Livestock Industry

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

Prof. Robin B. Gasser, University of Melbourne

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 & 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. Sang Yup Lee, Director, Bioinformatics Research Center, KAIST (Korea)
Prof. Peter Lindsay, School of Information Technology & Electrical Engineering, The University of Queensland; and Director, ARC Centre in Complex Systems
Dr Alex Loukas, Queensland Institute of Medical Research
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
- Prof. Paul Sternberg, California Institute of Technology, Pasadena (USA)
- Prof. Jenny Stow, Institute for Molecular Bioscience, The University of Queensland
- Dr Anne Trefethen, Executive Director, Oxford e-Research Centre (UK)
- Prof. Ah Chung Tsoi, Vice-President, Research & Institutional Advancement, Hong Kong Baptist University (Hong Kong)
- Dr Mark Wilkinson, Department of Medical Genetics, University of British Columbia and St Paul’s Hospital (Canada)
Prof. Janet Wiles, School of Information Technology & Electrical Engineering, The University of Queensland
Prof. Michael Zuker, Rensselaer Polytechnic Institute (USA)
Vision
The ARC Centre of Excellence in Bioinformatics is a multi-nodal interdisciplinary Centre that applies genome-scale bioinformatics, computational science and advanced data technologies to empirical data on cellular structure and function, yielding a deep understanding of the mammalian cell as an information system.

Mission
The ARC Centre of Excellence in Bioinformatics is committed to research excellence, advanced education and training, interdisciplinarity, improved access to skills, tools and facilities, and building critical mass and national focus for cellular and ‘omic’ bioinformatics in Australia.

Aims
Within this mission, we aim to:

• Build a powerful new bioinformatics that integrates insights and key developments in mathematics, statistics, computing science, information technology, genomics and molecular cell biology.
• Mutually enrich biomolecular research, mathematics, computer science and IT.
• Apply this new bioinformatics to make innovative use of advanced biomolecular technologies, with the result of illuminating the development and differentiation of the mammalian cell.
• Play a leading role in the national and international bioinformatics agenda and in the generation of human capital, enhancing critical mass and national focus in bioinformatics.

Objectives
More specifically, our objectives for the extension period will be to:

• Build on the capabilities and interdisciplinary perspectives we have developed so far.
• Use empirical data from advanced microarray and tomographic technologies to construct quantitative dynamic models of selected cellular regulatory networks and subcellular systems.
• Develop advanced algorithmic, statistical and computational methods and knowledge engineering systems, and apply them to extend, refine and inter-relate these models.
• Deliver advanced prototypes of our integrated data and visual framework for hypothesis-testing, validation and discovery, the Visible Cell™, and use them to understand more fundamentally the mammalian cell as an information system.

Capabilities — core technologies and infrastructure
The integrated research programs of the ARC Centre of Excellence in Bioinformatics are underpinned by established strengths in mathematics, computing science, information technologies, cell and structural biology, 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
• Advanced microscopic, tomographic and imaging technologies in cell biology
• High-throughput microarray-based genomic technologies
• Multiple interfaces to empirical “wet” biology
• Remote interaction via AccessGrid
Research Programs

Research programs highlights
The Centre’s research in 2007 was delivered through programs in experimental phenomics, modelling of cellular networks, and data integration and visualisation. Each of these programs was organised and managed as a set of projects. In the following pages we describe the aims of the 13 projects we managed during 2007, identify their personnel, and report their achievements during 2007.

2007 brought more than the usual amount of project realignment to our research portfolio. Several projects had been scoped to conclude in 2007. A few projects, or their infrastructural components, were transferred to the new Queensland Facility for Advanced Bioinformatics (QFAB). We seconded several ACB staff to QFAB to ensure continuity of infrastructure provision to our projects, and at the same time to help QFAB start up its operations and coordinate support across its partner projects; these staff will be drawn back into ACB as operational requirements allow. Certain other activities in the remaining projects were discontinued to allow a smoother transition to our Centre’s Phase-Two research programs.

Detailed planning for 2008 and beyond got underway on 24 April 2007, when all Chief Investigators in the extended Centre met in Brisbane. CIs and some postdoctoral researchers in our new Programs 2 and 4 met in Newcastle in June and in Brisbane in September, while CIs, postdocs and infrastructure staff in Programs 1 and 3 met in Brisbane in August 2007. As well, the second day of our 2007 Annual All-Hands Meeting (27 November 2007) was set aside for research planning. Our research portfolio for 2008 is outlined in the box on the right, while our detailed activity plans are set out in a separate (non-public) document that will be submitted to ARC concurrently with this Annual Report.

In Program 1, we applied high-throughput experimental methods to develop and test key hypotheses about the control of developmental processes in mammalian cells, and generated data required to model cellular and molecular-interaction networks. We used GFP and epitope tagging, high-throughput reporter screens for small molecule bioactives, 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 with Dr Brad Marsh and the Advanced Cryo-Electron Microscopy Laboratory – previously a node of the NANO-MNRF and now a peak node of the NCRIS-funded Australian Microscopy & Microanalysis Research Facility (AMMRF) – we have had unique access to high-resolution 3D electron tomography data on mammalian cells, including 3D reconstructions of whole mammalian cells at ~15-20 nm resolution. Activity in this program provided gene and protein networks for construction of computational models, and gave us capability to validate predictions against observed phenotype. Phenomic data also underlay our computational models of complex cellular structures, including dynamic spatial relationships among vesicles, lipid rafts and tubules during endocytosis.

The objective of Program 2 was 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 developed efficient algorithms, applied advanced methods from statistics and machine learning, integrated expertise in network representation and analysis, identified orthologs at unparalleled resolution, and made extensive use of high-performance computing.

In 2007, Program 3 began to deliver an innovative prototype data, modelling and visualisation environment, the Visible Cell™, that allows us to 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 required us to synthesise large quantities of diverse, multi-scale data including images, molecular structures and mathematical simulations; in this we worked closely with IMB and, during the second half of the year, with QFAB, the BioMANTA project, and the European QosCosGRID consortium.

Research programs 2008-2010

Program 1 – The Visible Cell™ linking genome and phenotype
1.1 Generation of subcellular localisation data in mammalian cells
1.2 Project and data management for Visible Cell™
1.3 Data analysis and knowledge discovery in very large biological network databases

Program 2 – Phenotype-informed discovery of networks and systems
2.1 Computational discovery and modelling of gene-regulatory networks
2.2 Differential expression of genes towards pathway discovery
2.3 Discovery and modelling of RNA-based regulatory networks

Program 3 – Modelling dynamic cellular processes
3.1 Mathematical modelling of genetic regulatory networks and biochemical pathways
3.2 Modelling and visualising biological processes in complex spatial environments

Program 4 – Algorithms for graphs and networks
4.1 Elaboration of the Vertex Cover problem
4.2 Optimisation-based approaches to large systems
4.3 External-memory algorithms for analysis of massive graphs
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, and Edinburgh Bioscience Research Centre, UK)

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 (2007):
Data collection for this project was largely completed in 2006. Our work in 2007 focused on data analysis and preparation of manuscripts based on research carried out in this project.
1.2 Computational cellular biology: combining computational prediction and experimental validation

Investigator:
Dr Rohan Teasdale
(The University of Queensland)

In collaboration with:
Dr Mikael Bodén
(RIKEN Genomic Sciences Center, Japan)
Prof. P. Carninci
(RIKEN Genomic Sciences Center, Japan)
Prof. Rob Parton
(The University of Queensland)
Dr H. Suzuki
(RIKEN Genomic Sciences Center, Japan)

Researchers:
Research Officers:
Dr Lynn Fink
(The University of Queensland)
Dr Nicholas Hamilton
(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 Josefine Sprenger
(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.

Achievements (2007):
(a) We combined updated human and mouse protein sets within the LOCATE database, and continued to populate it with experimental and literature-based subcellular localisation data.
(b) We defined organelle proteomes for network-based analysis. Definition of the nuclear proteome was based on our Experimental Location Proteomics Platform (ELPP) output from 2006, and its network analysis formed part of the RIKEN FANTOM4/Genome Network Project. The endosome proteome will form the basis to develop high-content screens within the Teasdale group.
(c) We established automatic image-capture and image-analysis protocols for the Experimental Location Proteomics Platform (ELPP) based on the BD Pathway high-content screening equipment purchased by IMB in 2006.
(d) We developed a subcellular localisation predictor for peroxisome-resident proteins (with M. Bodén, ITEE) and used it to identify putative peroxisome proteins for experimental validation. A paper arising from this work appeared in *Proteins: Structure, Function and Bioinformatics.*
(e) In collaboration with the group of Prof. Jenny Stow (IMB), Object Based Co-localisation (OBCOL) methodology and software has been developed and applied to 3D cell data. OBCOL allows the automated selection of all organelles, such as endosomes, from whole-cell 3D co-localisation imaging, enabling statistical calculations to be performed on each discrete structure. In this way, subtle changes in co-localisation under differing experimental conditions can be determined that were not detectable by whole-image co-localisation analysis. Manuscripts on methodology and on experimental applications are in preparation. As part of the OBCOL project, novel techniques utilising data dimension reduction have been developed to enable the co-localisation data generated to be visualised and compared across experiments.
(f) The iCluster high-throughput image visualisation and clustering tool was developed, and a paper outlining the principles appeared in print. iCluster enables large (~1500) bio-image sets to be visualised simultaneously and manipulated in 3D. Images are placed in 3D in such a way that visually similar images are spatially close, allowing easy comparison and differentiation of high-throughput image sets.
(g) Prof. Robert Murphy (Carnegie Mellon), a leader in the field of automated subcellular image classification, visited the ACB in December. An agreement was reached to exchange high-throughput microscopy data sets and analysis methodologies.
(h) Preliminary work has gone into integration of the ideas of the ASPiC and iCluster methodologies into the Visible Cell™ project. This will enable content-based querying, clustering and eventually visualisation of large image sets within the Visible Cell™. By allowing combinations of ontological and content-based searches, the range of inference possible will be significantly increased.
(i) The DomainDraw protein domain visualisation tool was published and was made publicly available both for local use and via a web interface. It is currently being used in the family of LOCATE databases, and has been requested and installed by a group in USA.
(j) In collaboration with the group of Dr Carol Wicking (IMB), a tool for quantifying cell recovery scratch tests has been developed. This enables the automated quantification of cell motility in minutes, rather than hours by hand-tracing images. A collaborative manuscript utilising this methodology is under review.
(k) A computational method for predicting helix tilt angles and helix-helix associations in membrane proteins was developed.
(l) A method for predicting protein-RNA and protein-DNA binding sites based on machine-learning and network theory was developed, and shows improved performance in comparison with existing approaches.
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 Assistant:**
Mr Radosav Pantelic  
(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 Nicholas Hamilton).
(e) Generate custom data in response to requirements of cellular modellers (Dr Nicholas Hamilton and Dr Jim Hanan).

**Achievements (2007):**
(a) A new set of some 5000 images divided into 10 subcellular localisations in HeLa cells has been created. The Automated Cell Phenotype Classification project (ASiC) was tested on the set, and was found to have a classification accuracy of 96%. This demonstrates that near-perfect classification is possible on images from fully automated microscopy image capture.
(b) The first image statistics specifically designed for subcellular localisation applications, threshold adjacency statistics (TAS), were published (*BMC Bioinformatics*), enabling an automated classification accuracy of up to 98% in the ASiC pipeline. The associated image sets were made publicly available via the LOCATE database.
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)
- Prof. Mark Ragan (University of Tennessee and ORNL)
- Prof. Mike Langston (University of Tennessee and ORNL)
- Prof. Rodney Scott (The University of Newcastle)

In collaboration with:
- Dr Timothy Bailey (The University of Queensland)
- Dr Regina Berretta (The University of Newcastle)
- Prof. Pablo Moscato (The University of Newcastle)
- Prof. Frank Dehn (Deakin University)
- Prof. Mark Ragan (The University of Tennessee)
- Prof. Rodney Scott (The University of Newcastle)

Research Associates:
- Dr Osvaldo Rosso (The University of Newcastle)
- Dr Fran Rosamond (The University of Newcastle)

PhD student:
- Mr Mario Instroza-Ponta (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 algorithms.

(c) Address open research issues of interest to the entire community of the ARC Centre, 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.

Achievements (2007):
(a) We developed a linear-size polynomial-time FPT kernelisation algorithm for the CLUSTER EDIT problem (presented at IPEC 2007).

(b) We developed a linear-size polynomial-time FPT kernelisation algorithm for the BOUNDED DEGREE DELETION problem, for any fixed bound (joint work with the FPT research group at Friedrich Schiller University, Jena; submitted to ICALP 2008).

(c) Jointly with the FPT research group at the University of Bergen, we developed an FPT algorithm based on poly(k) kernelisation for the CLUSTERING WITH UNCERTAINTY problem, parameterised jointly with the number of nodes in uncertain relations. A manuscript is in preparation.

(d) We developed a quadratic kernel FPT algorithm for the CONVEX RECOLORING problem (presented at COCOCOON 2007).

(e) We carried out parameterised complexity analysis of the GRAPH MOTIF problem for connected motifs (presented at ICALP 2007).

(f) We introduced into the field of bioinformatics new quantifiers of gene expression profiles based on Shannon entropy and statistical complexity ideas. In particular, application of these quantifiers to public-domain datasets (as well as to proprietary datasets at the Newcastle node) allowed us to explore their correlation with cancer progression. Two pilot studies were finalised, one on prostate cancer and the other on melanoma, both using microarray datasets. Using the mathematical formalism of the (\((\alpha,\beta)\)-k-FEATURE set (introduced by Cotta, Sloper & Moscato in 2004), we were able to find the sets of genes that characterise transitions of entropy and statistical complexity, and to correlate this with literature in prostate cancer and melanoma.

(g) Several publications had as a lead theme the identification of patterns in time series from a biological motivation. Viscoelastic properties of erythrocytes were studied from the perspective of a bounded correlated random walk (Brownian motion) based on the assumption that diffractometric data involve both deterministic and stochastic components. Photometric readings were obtained by ektacytometry over several millions of shear-elongated cells, using a home-made device called an erythrodeformeter. The results suggest that samples from healthy donors are intrinsically unpredictable (ordinary Brownian motion), while samples from \(\beta\)-thalassemia patients exhibit not only a great sensitivity to initial conditions (fractional Brownian motion) but also chaotic behaviour.

These results allow us to claim that we have linked nonlinear tools with clinical aspects of the rheological properties of erythrocytes. A paper (Korol, Foresto & Rosso) appeared in Physica A (2007).
(h) Chaotic systems share with stochastic processes several properties that make them almost indistinguishable by traditional approaches. We introduced a representation space, to be called the complexity-entropy causality plane, with horizontal and vertical axes that are suitable functions of the pertinent probability distribution (the entropy of the system, and an appropriate statistical complexity measure, respectively). These two functions were evaluated using the method of Bandt & Pompe (Phys. Rev. Lett. 2002) to assign a probability distribution function to the time series generated by the system. Several well-known, model-generated time series, usually regarded as being of either stochastic or chaotic nature, were analysed to illustrate the approach. The main achievement of this work is the possibility of clearly distinguishing between them in our representation space, something that is difficult otherwise. A paper (Rosso et al.) appeared in Phys. Rev. Lett. (2007).

(i) Using appropriate information theory quantifiers (e.g. normalised Shannon entropy and the Martin-Plastino-Rosso intensive statistical complexity measure), we revisited the characterization of Gaussian self-similar stochastic processes from a Bandt-Pompe viewpoint. We showed that the ensuing approach exhibits considerable advantages with respect to other treatments. In particular, values of the quantifiers present a clear gap in the transition between the continuous processes and their associated noise. A paper (Rosso et al.) appeared in Phys. Rev. E (2007).

(j) Two information theory-based indicators were used to measure the goodness of two encryption schemes commonly used within the context of chaotic communications. In particular, we have shown that the computation of the normalised Shannon entropy and the MPR statistical complexity measure for different chaotic laser signals can lead to statistically significant criteria for assessing the quality of several encryption techniques. The proposed measures allow, in some cases, to detect the presence of a message embedded within a chaotic carrier. They also reveal that the chaos modulation scheme is more reliable from the statistical point of view, when compared with chaos shift keying. A paper (Rosso et al.) appeared in Phys. Lett. A (2008).

(k) We explored deviations from efficiency in the returns and volatility returns of Latin-American market indices. Two different approaches were considered. The dynamics of the Hurst exponent was obtained via a wavelet rolling sample approach, quantifying the degree of long memory exhibited by the stock market indices under analysis. On the other hand, the Tsallis q entropic index was measured in order to take into account the deviations from the Gaussian hypothesis. Different dynamic rankings of inefficiency were obtained, each of them contemplating a different source of inefficiency. Comparing with results obtained for a developed country (USA), we confirmed a similar degree of long-range dependence for our emerging markets. Moreover, we showed that the inefficiency in the Latin-American countries comes principally from the non-Gaussian form of the probability distributions. A paper (Zunino et al.) appeared in Eur. Phys. J. B (2007).

(l) Using a mathematical model based on the QUADRATIC ASSIGNMENT problem, we developed software to visualise large numbers of clusters of gene expression profiles. We validated the methodology using a publicly available yeast cell-cycle dataset. Our method allows the set of gene profiles to be partitioned into clusters that guarantee that the k-closest profile, for any given profile, belongs to the same cluster. In addition, two optimisation heuristics (based on memetic algorithms) work together to optimise the layout of genes and the layout of clusters. The aim is to annotate putative functions to genes for which we have little information, by looking at the Gene Ontology annotation of other genes in the cluster. The methodology was validated in other problem domains, and is currently being investigated as a tool to understand the different metabolism of tumour cells in different cancer subtypes. A paper (Inostroza-Ponta et al.) appeared in Lectures Notes in Computer Science (2007).
2.2 Computational discovery of gene-regulatory motifs and motif clusters

**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 (2007):**

(a) Tom Whitington carried out preliminary work toward developing models of tissue-specific regulation of genes involved in erythropoiesis, using bioinformatics tools to predict the targets of EKLF and GATA-1. The results remain incomplete because current tools cannot accurately predict transcription factor binding sites. Collaborators in Andrew Perkins’s group worked to develop a satisfactory chromatin immunoprecipitation (ChIP) experiment to identify EKLF targets; this work is ongoing. Further activity towards this goal by Tom Whitington remains on-hold, pending the ChIP results.

(b) Tom Whitington and Mikael Bodén developed a bioinformatics pipeline to help biologists develop tissue-specific gene regulation models. This pipeline identifies which transcription factor binding profiles are significantly over-represented in groups of genes that share functional annotation in the Genome Ontology (GO) database. We hope that this approach will provide additional support for predictions of tissue-specific targets of transcription factors. A manuscript is currently in preparation.

(c) Emma Redhead completed the development and testing of her DEME software (Discriminatively Enhanced Motif Elicitation) that uses both positive and negative data for motif discovery, and demonstrated its superior utility for discovering motifs in protein and DNA sequences in certain settings. In particular, this appears to be a very useful approach for identifying discriminating sequence features in protein families. This work resulted in a manuscript in *BMC Bioinformatics*. The software is freely downloadable for academic use via the ACB website at http://bioinformatics.org.au/deme.

(d) Denis Bauer produced software for improved prediction of transcription factor binding sites utilising protein-protein interaction data. This software and the resulting predictions were provided to the RIKEN consortium, and a manuscript is currently in preparation.

(e) In conjunction with the Noble group at the University of Washington and the Wilfred Li group at UC San Diego, we developed next-generation webservers for MEME and several additional bioinformatics programs. These webservers provide many advanced features including the caching of all inputs and outputs at the remote server. This facilitates “pipelining” of results from one application to another. It also eliminates many of the problems associated with sending large results files by email.

(f) We developed and deployed a webservice for the motif comparison tool Tomtom. This tool was also integrated with MEME and GLAM2 (see below) so users can conveniently compare their newly discovered sequence motifs with existing databases of known motifs, including Transfac and JASPAR. A paper describing Tomtom was published in *Genome Biology*.

(g) Visiting researcher Osamu Maruyama completed development and testing of his phylogenetic motif discovery tool. A collaborative manuscript is in the final stages of preparation.
(h) In collaboration with William Noble, we continued work toward an algorithm for ortholog-based motif scanning algorithm. The surprisingly poor performance of ortholog-based motif scanning in the well-studied yeast case led to a project with John Hawkins to study the theoretical power of such methods. We discovered that all extant tools fail to achieve anything approaching the theoretical power of the method because one or more of the assumptions about nature made by the methods is erroneous. We are now studying the reasons for this failure of the tools in practice. We anticipate that a manuscript describing this work will appear in the proceedings of RECOMB 2008.

(i) Collaborative work with former postdoctoral fellow Martin Frith on the GLAM2 algorithm for discovery of gap-containing motifs led to a manuscript which has been accepted for publication by *PLoS Computational Biology*. The GLAM2 software is available for free download from the ACB website (http://bioinformatics.org.au/ glam2). In addition, we integrated the GLAM2 software into the MEME Suite package, and created web servers for GLAM2 and GLAM2SCAN. As mentioned above, the output of GLAM2SCAN was integrated with the Tomtom webserver for integrated analysis.
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. Jangsun Baek  
  (Chonnam National University, Korea)
- Prof. Kim-Anh Do  
  (MD Anderson Cancer Center, University of Texas)
- Dr Richard Kerr  
  (MMi Genomics)
- Dr Thriyambakam Krishnan  
  (Systat Bangalore, India)
- Dr Angus Ng  
  (The University of Queensland)
- Prof. Erick Suarez Perez  
  (University of Puerto Rico)
- Dr Sam Wang  
  (The University of Queensland)

**Researchers:**
- Research Associate: Dr Ian Wood  
  (The University of Queensland)
- Research Assistant: Mr Lloyd Flack  
  (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 (2007):**
(a) In 2007 we continued with our research on problems associated with the detection of differentially expressed genes from tissue samples under various experimental designs. In particular, we developed and investigated extensions of the methodology to allow for the fact that not all the genes are independently distributed. Also, we continued with the development of computer programs to produce user-friendly software for these problems, for example in R.

(b) We continued with a study of multiple-comparison procedures and their application to genomics beyond those problems considered in (a). A manuscript is being prepared on this work.

(c) A manuscript on image analysis of expression data using two-channel cDNA arrays appeared in *Bioinformatics*.

(d) We continued with work on the use of gene-expression data in the construction of genetic regulatory networks (GRNs).

(e) We investigated selection biases that occur in visualising and forming classifiers and other statistical rules on the basis of small subsets of genes selected in some optimal (non-random) way from a much larger set. Two manuscripts on this work have been drafted for publication.

(f) We essentially completed writing of new material for the second edition of the Wiley monograph on the EM algorithm and its extensions, after a decision was taken in 2007 to expand the first draft of the second edition completed in 2006.

(g) We studied further the problem of clustering microarray data, and two book chapters on this topic were completed; a third chapter is in preparation.

(h) We collaborated with Project 2.6 to characterise the magnitude of the state space available for randomised controls in the prediction of miRNA targets in the human and mouse transcriptomes.
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 Manuel Barrio  
(Valladolid University, Spain)  
Dr Pamela Burrage  
(The University of Queensland)  
Dr Margherita Carletti  
(Università di Urbino, Italy)  
Dr Grant Lythe  
(University of Leeds, UK)  
Mr Dan Nicolau Jr  
(Oxford University)

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

PhD students:  
Mr Alhadi Bustamam  
(The University of Queensland)  
Mr Shev MacNamara  
(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 (2007):
(a) We modelled and simulated the dynamics of the Hes1 molecular clock in mouse. This was collaborative work with Kevin Burrage (UQ) and researchers in Spain (M. Barrio) and co-workers at the University of Queensland. The work was reported in PLoS Computational Biology.
(b) In collaboration with Andrés Leier and Tatiana Marquez Lago (both now at ETH Zürich), we extended our ideas on the Hes1 molecular clock and modelled and simulated the dynamics of the Her1 and Her7 complex in somite formation in zebrafish. This work was presented at the International Conference on Computational Science 2007 in Beijing, and an article will appear in a special journal issue in 2008 associated with this conference.
(c) We developed and implemented multiscale 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. We employed these ideas to significant biological problems including the kinetics of the MAP kinase cascade. Variants of this work appeared in three journal articles, in particular a special issue on the 150th anniversary of Markov, and a special issue of International Journal of Computational Science. This work was carried out by Kevin Burrage and PhD student Shev MacNamara at UQ, and Dr Roger Sidje, who is now based at the University of Minnesota.
(d) Kevin Burrage and Jiangning Song (now at Kyoto) developed innovative methods for predicting disulfide connectivity from protein sequence using multiple sequence feature vectors and secondary structure. This work appeared in Bioinformatics in 2007. A web server was provided to researchers for these predictions.
(e) In collaboration with Ian Lenane (UQ) and Grant Lythe (Leeds), we developed new numerical methods for second-order stochastic equations. This work appeared in SIAM Journal of Scientific Computing.
(f) We developed some algorithms for solving fractional differential equations that arise when modelling anomalous diffusion on the plasma membrane. This work, a collaboration between Kevin Burrage and co-workers in China, appeared in two articles in Journal of Applied Mathematics and Computation.
(g) Kevin Burrage and student Tim Rudge (now in London) developed new stochastic models describing how the effects of intrinsic and extrinsic noise can accelerate and stabilise juxtacrine pattern formation. This work has appeared in Bulletin of Mathematical Biology.
(h) We developed a new project on large-scale implementations for Markov clustering involving Kevin Burrage, Mark Ragan, Shoaib Sehgal, Simon Wong and PhD student Alhadi Bustamam. The first application, to ortholog determination in mammalian genomes (Project 2.6), has been submitted for publication.
2.5 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:**

**Senior Research Fellow:**
Dr Mikael Bodén  
(The University of Queensland)

**Postdoctoral Research Fellow:**
Dr James Watson  
(The University of Queensland)

**MIT (Masters by coursework) student:**
Mr Andres Sanin  
(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 (2007):**
In 2007, tools developed in earlier stages of this project were applied to a range of genetic regulatory networks:
(a) The Artificial Genome Algorithm was analysed using the distributed simulation kit D’Pipeline, and the results were presented at Information Processing for Cells and Tissues (IPCAT’07) in Oxford.
(b) Multi-dimensional analysis tools were used to study robustness and dynamics in the state space structure of the *Drosophila melanogaster* segment polarity network and the *Saccharomyces cerevisiae* cell-cycle network; a paper was submitted for publication.
(c) A tool for distributed parameter sweeps was developed and tested on the Artificial Genome Model and applied to transcription site prediction in *Bacillus subtilis*.
(d) This project came to its planned conclusion at the end of Phase one of the Centre. Some applications continue and will be reported in future through the Centre as appropriate.
2.6 Ortholog mapping of prokaryotic and eukaryotic genomes

**Investigator:**
Prof. Mark Ragan  
(The University of Queensland)

**In collaboration with:**
Dr Robert Beiko  
(Dalhousie University, Canada)
Prof. Mike Fellows  
(University of Newcastle)
Dr. Jonathan Keith  
(Queensland University of Technology)
Dr István Miklós  
(Hungarian Academy of Sciences)
Dr Alexey Murzin  
(MRC Laboratory of Molecular Biology, Cambridge UK)
Dr Fran Rosamond  
(University of Newcastle)

**Researchers:**

**Research Associates:**
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 Chang Jin Shin  
(The University of Queensland)

**Intern:**
Mr Arnab Mandal  
(Indian Institute of Technology Kharagpur)

**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).

(a) Develop our automated high-throughput bioinformatic pipeline for the identification of protein family clusters, optimal multiple sequence alignment, inference of statistically sound phylogenetic trees, and rigorous comparison of subtrees. In particular, implement and test algorithms and tools that allow the automated recognition and analysis of arbitrarily or self-defined sub- and super-genic regions that have been affected by genetic recombination or LGT.

(b) Apply this pipeline to genome sequences of specific prokaryotes to identify orthologous and paralogous gene families, and elucidate pattern and mode of genome and gene-family evolution. Develop and test hypotheses about the frequency, mechanisms, patterns and biological significance of genetic recombination and LGT in specified taxa, particularly those of pathogenic bacteria.

(c) Extend the application of this pipeline to additional data domains, particularly complex eukaryotic genomes. Generate a novel database of orthologous and paralogous gene regions (defined at high resolution) among vertebrate genomes including those of human, mouse, and teleosts. Develop and test hypotheses about the frequency, mechanisms, patterns and functional significance of gene duplication in these taxa.

(d) Interrelate the database of orthologous and paralogous gene regions with data on protein–protein and other biomolecular interactions in mammalian cells. Use known and inferred biological relationships (*e.g.* ortholog relationships) in quality control of protein–protein interaction data. Develop and test hypotheses relating evolutionary dynamics of gene families and genomes to structure and biological function of molecular interaction networks.

**Achievements (2007):**

(a) In collaboration with Robert Beiko, we applied our automated high-throughput bioinformatic pipeline to characterise vertical and lateral gene transfer in genus *Staphylococcus*. This was done separately for a dataset of orthologs in 13 *Staphylococcus* genomes (PhD student Cheong Xin Chan) and for a dataset that includes paralogs although for only 8 genomes (former honours student Dave Tang). The former manuscript was lodged on arXiv.org in September 2007, and submitted for journal publication in February 2008.

(b) In collaboration with Robert Beiko, Cheong Xin Chan completed the assessment of a two-phase approach to discovery of genetically recombined regions and recombination breakpoints using simulated data, and implemented the code on the APAC national supercomputing facility. A manuscript describing the method was invited for journal publication. In further collaboration with Aaron Darling, we applied our approach to the 144-genome alignment sets examined earlier (Beiko et al., *PNAS* 2005), allowing us to map recombination breakpoints onto genes known to be of vertical or lateral ancestry. By further mapping SCOP structural domains onto these families, we set up the first-ever test of the hypothesis that (genomic regions encoding) protein structural domains are the units of lateral transfer. The results are tantalising but have proven to be statistically slippery. A collaboration has been struck.
(c) Aaron Darling, in collaboration with István Miklós, has developed Bayesian models to infer the most-probable history of genome rearrangements in genus *Yersinia*. The results, which one reviewer termed “the future of comparative genomics”, strongly imply that replichore balance is under active selection. A manuscript is under second-stage review at *PLoS Genetics*.

(d) Simon Wong completed the design of algorithms for fine-scale ortholog mapping among large eukaryotic genomes, and their implementation on the APAC national facility. The results can be abstracted as a graph with 1.05 million vertices and about 280 million edges. We developed new approaches to allow recovery of several tens of thousands of cliques or near-cliques, depending on the edge probability threshold. A manuscript was submitted to ISMB 2008. We collaborated with Mike Fellows and others to define this as the *EDGe DISJOINT CLIQUE COVER* problem; a quadratic kernel FPT algorithm was developed, and a manuscript submitted to CATS 2008. Problems of this size will become increasingly common in systems biology. We collaborated with Alhadi Bustamam and Kevin Burrage (Project 2.4) to develop efficient parallel computing structures for this problem, resulting in a second manuscript submitted to ISMB 2008.

(e) Karin Kassahn implemented the *iADHoRe* software (*Semillion et al. 2008*) on the APAC national facility and extended components of our eukaryotic pipeline to map regions of genomic duplication in five teleost fish. Owing to unanticipated limitations in the ENSEMBL gene trees, we have been forced to re-infer trees for several thousand eukaryotic gene families. Extensive tests give us confidence that we are distinguishing gene copies that descend from the whole-genome duplication event at the base of the teleost lineage from those that originate in later segmental duplications. During the first half of 2008 we will submit at least one manuscript relating patterns of ancestral genome duplication and loss to the appearance of new functions (neofunctionalisation) in fish genomes. In parallel work, student Vinh Dang used text-mining techniques to construct the first comprehensive database of haploinsufficient genes in human, including descriptions of related medical conditions. A manuscript was invited for review (February 2008).

(f) Chang Jin Shin completed the construction of a highly curated protein-protein interaction database for human and mouse that includes annotation of underlying experimental evidence, and subcellular localisation. A software module allows navigation of Gene Ontology hierarchies. These resources have allowed us to explore the hypothesis that interacting proteins must be co-localised in the cell, and that orthology information (“interologs”) can improve the inference of protein-protein interaction. Two manuscripts are in advanced drafts.
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:
Ms Kimberly Begley
(Griffith University, and APAC Grid Bioinformatics Program)
Dr Christopher Bouton
(Pfizer Research Technology Center)
Dr David Hansen
(QFAB)
Prof. Jane Hunter
(The University of Queensland)
Dr Ralf Muhlberger
(The University of Queensland)
Prof. Bernard Pailthorpe
(The University of Queensland)
Dr Michael Pheasant
(The University of Queensland)

Researchers:

Research Associates:
Dr Melissa Davis
(The University of Queensland)
Ms Helen Zi Huang
(The University of Queensland)

MSc student:
Ms Jooyoung Choi
(The University of Queensland)

ACB staff:
Mr Oliver Cairncross
(ARC Centre in Bioinformatics)
Ms Mhairi Marshall
(ARC Centre in Bioinformatics)
Mr David Wood
(ARC Centre in Bioinformatics, & QFAB)

Intern:
Himanshu Prakash
(Indian Institute of Technology, Kanpur)

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, medium- and high-resolution cell 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, specialised mark-up languages, and Semantic Web technologies.
(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) In collaboration with the Queensland Facility for Advanced Bioinformatics (QFAB) and national research infrastructure programs including NCRIS 5.16, Grid-enable key research databases and selected data within the Centre.

Achievements (2007):
(a) SRB has been operational within ACB since mid-2006. During 2007, management of this infrastructure was transferred to QFAB. SRS scripts were implemented and mirrors were established within SRB for a number of external databases including GenBank, RefSeq, UniGene, UniProt, Interpro, Pfam, Rfam, PDB, GO, KEGG, Reactome and SRS itself. These database sources were made accessible via the SRB webservice mySRB (https://www.sdsc.edu/srb/mySRB/mySrb331all.shtml) and via the Windows desktop client inQ (http://www.qfab.org.au, follow links to Databases & Tools / SRB). We undertook collaboration with researchers at James Cook University who are developing a new multi-platform desktop client HERMES, an inQ-like but multi-platform client that furthermore supports Shibboleth and connects to local servers via LDAP.
By Q2 2008 we hope to provide SRS and the UCSC genome data via SRB to researchers across Australia for use from their desktop. The link with the SRB client at Griffith University was delayed for reasons beyond our control; connectivity is anticipated during Q1 2008.
(b) Issues remain around stability of Cell Centered Database (CCdB) and its installation in our alpha site. The UCSD development team at the US National Center for Microscopy and Imaging Research (NCMIR) at UCSD was given access to ACB servers in Q3 2007, but installation in our environment remains problematic. As a consequence, little progress could be made on development of the Visible Cell™ data store.
(c) As indicated in (a) above, SRS 8.2 was implemented and is production. In 2007 we brought on-line a batch process that integrates data from the SRS store in SRB into the Visible Cell™. Work is underway on interfacing the Visible Cell™ with our SRS federation via Web services, and is on track to be operational by Q2 2008. A Web services interface for the LOCATE database (Project 1.2) was implemented successfully, uniquely linking in-house subcellular localisation data into our SRS federation.

Research Programs

2007 ANNUAL REPORT
ARC CENTRE OF EXCELLENCE IN BIOINFORMATICS
35
(d) In collaboration with Michael Pheasant and QFab, the UCSC Genome Browser mirror was brought on-line as an open resource. During 2007 the size of the underlying data exceeded a threshold beyond which the existing approach to data synchronisation was no longer adequate; we are working with UCSC and hope to have a resolution by Q2 2008.

(e) We continued to investigate methods for summarisation of protein-structure data in 3D and for reduction of dimensionality, focusing on (1) comparative testing of BCS versus other dimensionality reduction methods, and LCM versus other data-summarisation methods, using large protein-structure datasets in order to establish a systematic understanding of the relationship between structure-similarity query-processing strategies and their performance and accuracy on 3D protein structure datasets, and (2) the impact of similarity threshold on query-processing performance in order to automate the tuning process for selecting values for key parameters used in indexing and query processing.

(f) As part of the BioMANTA collaboration with Chris Bouton at Pfizer Research Technology Center (Cambridge, Massachusetts USA) we developed an approach to integrate hierarchical controlled vocabularies with OWL ontology, using protein-protein interactions as a case study. A paper was accepted at APBC-2008. Also within BioMANTA, in collaboration with the group of Jane Hunter, we established an RDF triple store for protein-protein interaction data and investigated how it scales to 100 million triples and beyond.

(g) We collaborated with researchers in Project 3.2.1 to design an ontology for subcellular localisation in the mammalian cell. Our SCL Ontology is already being used within a semi-supervised approach to inferring networks based on data extracted from independent data sources (here gene expression, subcellular localisation, and ortholog mapping).

(h) In order to bring small-molecule data into the same reasoning framework as cell images and protein-protein interaction networks, we initiated a project to represent small-molecule data using Semantic Web technologies. A survey of how small-molecule data are represented in online data sources was completed, and a manuscript is underway.
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)
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 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 (2007):
(a) Up to now, this project utilised data from expression microarrays for analysis of gene expression, under the assumption that the amount of an mRNA present in a cell approximates the expression level of the corresponding protein. However, the role of non-coding RNA in post-transcriptional regulation of gene expression has recently become more widely appreciated. Accordingly, as part of the transition to Phase Two of the Centre, we began to shift the focus of this project to non-coding RNAs, especially the genes that are targeted by these RNAs in 2008. Given that a single miRNA can target multiple mRNAs, and a single mRNA can be targeted by different miRNAs, this new focus will more naturally yield information on networks of regulatory control of gene expression.
(b) In 2007 we developed a method to predict miRNAs targeted by miRNAs. Several miRNA target-prediction methods having been proposed, most of them based on complementary binding in the 3’ UTR, cross-species conservation of sites, and minimisation of free energy. Our method is based on bulge-tolerant complementary binding sites. We applied this method not only to 3’ UTRs but also to predict complementary sites in the 5’ UTRs. We will continue this work in 2008.
(c) We initiated a project to group miRNAs according to their functions. To this end we developed and implemented a novel algorithm based on decomposition of sub-sequences.
(d) We implemented an index mechanism for DNA sequences and amino acid sequences. BLAST is the most popular sequence matching algorithm, but has shortcomings in dealing with bulge-tolerant sequences. We constructed our algorithm to complement this deficiency in BLAST.
(e) We applied our interactive 3D visualisation model to visualise very large biological sequences data, in which end-users can select dimensions in which to display the data distribution. We applied this approach to other gene-expression datasets.
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:
Mr. Oliver Cairncross
(Arc Centre in Bioinformatics)

In collaboration with:
Dr. Ben Hankamer
(The University of Queensland)
Prof. Jane Hunter
(The University of Queensland)
Dr. Rohan Teasdale
(The University of Queensland)

Researchers:
Principal Research Fellow:
Dr. Jim Hanan
(The University of Queensland, jointly with ACMC, ACCS and ACLR)

Research Associates:
Dr. Nicholas Hamilton
(The University of Queensland)
Dr. Muhammad Shoaib Sehgal
(The University of Queensland)

PhD students:
Mr. Andrew Noske
(The University of Queensland)
Mr. Peter van de Heide
(The University of Queensland)

ACB staff:
Mr. Matthew Bryant
(Arc Centre in Bioinformatics)
Mr. Oliver Cairncross
(Arc Centre in Bioinformatics)
Mr. Tim McComb
(Arc Centre in Bioinformatics)
Ms. Mhairi Marshall
(Arc Centre in Bioinformatics)
Mr. Tim Sullivan
(Arc Centre in Bioinformatics, & QFAB)
Mr. David Wood
(Arc Centre in Bioinformatics, & QFAB)

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 (2007):
(a) Considerable work has been devoted to Visible Cell™ architecture. The internal data structure has been redefined so that it is now an ontology. Integrating data in the Visible Cell™ with other ontological data facilitated by creating mappings between ontologies. For instance, when mapped to Gene Ontology, queries can be performed with respect to both cellular reconstruction data and the Gene Ontology hierarchy. In 2007 we demonstrated the use of SPARQL in querying over these datasets. This redesign addresses or obviates several specific activities set out in our Activity Plan for 2007.

(b) We completed the implementation of procedures and code to visualise and query pathway data in KEGG and Reactome. With KEGG, the system is now able to retrieve and display specific sub-cellular components at each annotated step of a metabolic pathway. This allows researchers to visualise the location of pathway activity when applied to a reconstructed cell. This was made possible by integrating spatial data (Brad Marsh) and protein-localisation data (Rohan Teasdale) with federated databases to build a mapping between sub-cellular components and metabolic pathways.

(c) With the redefinition of the Visible Cell™ internal data structure as an ontology, we postponed public release of the software. Revision 3 (underway at the end of 2007) will provide access to all cell-reconstruction objects present in the data and covered by public ontologies, not only a subset.

(d) The version of the Visible Cell™ software developed during 2007 supports interaction with external modelling programs via a stable API. Using this API, an application is able to navigate the data stored in Visible Cell™ in a context-sensitive manner, and can load, manipulate and save it. Work has also commenced to enable the Visible Cell™ to initiate interactions with generic external applications, the results of which will be able to be integrated with its ontology-based data.
(e) Using the integration technology developed (above), Visible Cell™ data has been extracted in real time by an external scientific modelling program to model vesicular behaviour in a stimulated pancreatic beta cell. External modelling software is now able to communicate with the Visible Cell™ directly over the network, rendering unnecessary the import and export of (potentially large) data files.

(f) The graphics engine of the Visible Cell™ visualiser has been changed from a scientific visualiser library to one more in-line with a games engine. This change has resulted in a substantial increase in graphics performance of the system, making it more capable for visualisation of and interactivity with an atlas of the mammalian cell.

(g) We continue to support and enrich the project’s public website (http://www.visiblecell.com) and the internal project wiki.
3.2.2 Modelling and visualisation of biological processes in complex spatial environments

Investigators:
Prof. Kevin Burrage  
(The University of Queensland)  
Dr Timothy Bailey  
(The University of Queensland)  
Dr Markus Hegland  
(The Australian National University)  
Dr Rohan Teasdale  
(The University of Queensland)

In collaboration with:
Prof. John Hancock  
(The University of Queensland)  
Mr Dan Nicolau Jr  
(Oxford University)  
Prof. Rob Parton  
(The University of Queensland)  
Prof. Jenny Stow  
(The University of Queensland)

Researchers:
Principal Research Fellow:  
Dr Jim Hanan  
(The University of Queensland, jointly with ACMC, ACCS & CILR)

Research Associates:  
Dr Nicholas Hamilton  
(The University of Queensland)  
Dr André Leier  
(ACMC, The University of Queensland)  
Dr Tatiana Marquez  
(ACMC, The University of Queensland)  
Dr Jiangning Song  
(ACMC, The University of Queensland)

PhD students:  
Mr Alhadi Bustamam  
(The University of Queensland)  
Mr Shev MacNamara  
(The University of Queensland)

Aims:
Many cellular processes involve situations where the number of interacting molecular species is small or modest, and/or the surrounding medium is spatially complex, leading to long mixing timescales. In such cases, continuity assumptions (such as in rate constants or diffusion coefficients) may lead to approximations that are not useful, that is, fail to abstract actual behaviour. In this project we will develop codes for the (initially Monte Carlo) simulation of diffusive cellular processes under such situations. We will then link these simulators to visualisation strategies, notably those being used in the Visible Cell™ project.

Achievements (2007):
(a) Kevin Burrage and John Hancock (UQ) with Dan Nicolau Jr (Oxford) implemented new stochastic simulations for modelling anomalous diffusion and exploring the role of lipid rafts in the plasma membrane. The work was reported in Journal of Biophysics.
(b) We were invited by Briefings in Bioinformatics to write a review article on modelling and simulation techniques for membrane biology (Briefings in Bioinformatics 8:234-244, 2007).
(c) In joint work with Tatiana Marquez Lago (now at ETH Zürich), we developed new multiscale algorithms for the spatial modelling of discrete chemical kinetics in a cell. These ideas were based on tau-leap temporal methods, and this work was the first to use these ideas in a spatial context. This work was published in the Journal of Chemical Physics.
(d) Shev MacNamara, Kevin Burrage (UQ) and Roger Sidje (now at Minnesota) developed new numerical methods for modelling receptor oligomerisation using the chemical master equation. This work was presented at the 8th International Congress on Industrial and Applied Mathematics (Zürich 2007) and appeared in the conference Proceedings.
(e) Jim Hanan (UQ) and co-workers coupled stochastic spatial simulators with the Visible Cell™ prototype (Project 3.2.1) using Cell Illustrator™ and an L-systems based spatial prototyping system. However, redevelopment of the Visible Cell™ API as part of Revision 3 now provides a more generic approach to interaction with external applications such as Cell Illustrator™.
(f) Kevin Burrage and Tim Sullivan (UQ) developed and implemented a qualitative model of the intracellular processes involved in the production and release of insulin by mammalian beta cells, and interfaced this model with the Visible Cell™ visualisation system. This work not only provides insight into insulin dynamics, but also has allowed us to explore issues associated with incorporating modelling into the Visible Cell™ environment. A manuscript is in preparation.
The year 2007 saw the startup of BioMANTA, a new collaborative systems biology research project involving research groups at the University of Queensland (the School of Information Technology and Electrical Engineering, the Institute for Molecular Biosciences, and the ARC Centre of Excellence in Bioinformatics) and Pfizer’s Global Research and Development Center of Emphasis in Computational Sciences in Cambridge, Massachusetts USA. The aim of the project is to create a semantic web resource for knowledge discovery in integrated biomedical data. Specifically, we hope that by inferring the connectivity of biological networks and modelling their activity, we will discover new directions for pharmaceutical investigation, and improve the success rate of the drug discovery process.

The Semantic Web, as envisioned by Tim Berners-Lee and the W3C, will be an extension of the current Web in which “information is given well-defined meaning, better enabling computers and people to work in cooperation”. The vast majority of information online is not, or at least not easily, human-readable. If these resources were properly structured both semantically and syntactically, it would be possible to apply powerful machine-inferencing approaches to ask questions not only quantitatively but even qualitatively different from those that can be posed today: for example, about all the G-protein-coupled receptors, or the functional nature of all pathways represented in the set of shortest paths between a certain class of drugs and, say, apoptosis.

At Pfizer the project was initiated by Chris Bouton, head of Integrative Data Mining at the Cambridge centre. His team consists of Victor Farutin, Michael Schaffer and Fred Jervis, who work on visualisation and network meta-analysis. Professor Jane Hunter (UQ ITEE) leads the semantic web research component of the project, with team members Andrew Newman, Imran Khan, and postdoc Yuan-Fang Li. Within ACB, Mark Ragan and postdoc Melissa Davis are developing ontological knowledge representations of biomolecular data, while Kevin Burrage and postdoc Shoaib Sehgal are developing approaches to infer currently unknown connectivity within networks. Our interdisciplinary teams maintain collaboration and coordination via a project wiki, Skype, email and, for the UQ groups, regular face-to-face interaction.

2007 was a great initial year for BioMANTA (www.biomanta.org). The modellers developed a comprehensive, if ambitious, approach to network inference based on protein-protein interaction data, subcellular localisation and orthology, and plan to extend it to expression data. We developed an approach for integrating OBO and OWL ontologies, and migrated several large public biomolecular data sources to RDF. The semantic web team implemented an RDF triple store and demonstrated its scalability to tens of millions of triples (although at least one further order of magnitude will be required). During the fourth quarter of 2007 we presented several conference posters and talks (including one via AccessGrid to Korea), and three peer-reviewed papers were accepted for international conferences. With our group now assembled and momentum building, we anticipate that 2008 will be even more productive.

Melissa Davis
ACB’s Key Performance Indicators (KPis) encompass the following areas:

1. Research findings and competitiveness
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 2007:

The ARC Centre in Bioinformatics again hosted and sponsored the annual Winter School in Mathematical and Computational Biology. Our co-sponsors were Queensland Cyber Infrastructure Foundation (QCIF), The MathWorks Australia, SGI, and the Institute for Molecular Bioscience (IMB). 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 and engineering. Our fourth Winter School attracted 190 participants from 35 institutions in four countries.

Throughout 2007 our Centre was involved in numerous ways with Bioinformatics Australia, an activity centre within Australia’s peak body in biotechnology, AusBiotech. ACB alumna Dr Lucia Santoso managed the Australian Bioinformatics Network for Bioinformatics Australia. Rohan Teasdale chaired the organising committee for the 2007 Bioinformatics Australia conference, held in conjunction with the annual AusBiotech conference. Phoebe Chen, Mark Ragan (vice-president) and Shoba Ranganathan were members of BA’s steering committee.

ACB hosted the second Student Symposium in Bioinformatics co-sponsored under terms of our MOU with Bioinformatics Institute New Zealand. Students and staff from all four of our geographical nodes participated.

2007 was a good year for publications, with 66 C1 journal articles (35 of which were co-authored with overseas collaborators) and 39 E1 conference papers.

Two next-generation DNA sequencing platforms were installed in the laboratory of Sean Grimmond under a strategic research collaboration with Applied Biosystems. Sean reported initial results from deep sequencing of mammalian transcriptional complexity.

The BioMANTA project, involving researchers in ACB, the School of ITEE at the University of Queensland, and Pfizer Research Technology Center (Cambridge, Massachusetts) got underway in early 2007. BioMANTA is exploring the application of Semantic Web technologies to integration of protein-protein interaction data and modelling of signalling networks in mammalian cells.

Two ACB researchers are partners in the QosCosGRID Project funded by the EU FP6 Information Society Technologies program. QosCosGRID links 12 universities, research institutes and companies in six European countries, Israel and Australia in middleware development and implementation of a computing grid testbed for the modelling and simulation of complex systems in several domain areas including bioscience.

On behalf of and jointly with UQ, IMBcom filed for registration of Visible Cell™ as an international trademark in China. We own relevant Web domain names including www.visiblecell.org and www.visiblecell.com.

ARC Federation Fellow Professor Kevin Burrage accepted a joint appointment as Professor of Computational Systems Biology at Oxford University, and will spend several months there each year.

Professor Mike Fellows, from our University of Newcastle node, took up a senior Humboldt Fellowship. During 2007 he spent extended periods with top European research centres in mathematics and computational biology.

Professor Geoff McLachlan accepted appointment to the ARC College of Experts.

Dr Sean Grimmond and Dr Rohan Teasdale took up NHMRC Senior Research Fellowships.

Queensland Facility for Advanced Bioinformatics recruited Mr Jeremy Barker as CEO and Dr Dominique Gorse as Technical Manager. QFAB established an Industrial Research Advisory Panel, enrolled commercial clients, and secured a partnership within a major research grant funded by the National Breast Cancer Foundation. Jeremy Barker was awarded a Winston Churchill Memorial Trust fellowship to study management best practice in the delivery of bioinformatics services in USA and UK.
1. Research findings and competitiveness

Publications

Quality of publications

ACB researchers continue to perform very well in peer-reviewed publication, indicating the innovative and forefront nature of our research. By the narrowest definition our core Investigators published 66 C1 journal articles in 2007; if we include papers arising via existing collaborations with investigators who will join ACB in 2008, the total rises to 87. This performance illustrates the impetus that ARC funding of our Centre (which became available to us in late 2004) has given bioinformatics research in Australia. Of our 66 core C1 journal articles, 31 (47%) appeared in journals with 2006 impact factor ≥3.0, 14 (21%) with iF ≥5.0, and 4 (6%) with iF ≥10.0. Our impact beyond bioinformatics per se is reflected by publications in prestigious journals including Current Opinion in Microbiology, Genome Biology (2), Nature, Nature Medicine, Nature Methods (2), Nucleic Acids Research (3), Physical Reviews E and Systematic Biology. 35 (53%) of C1 journal articles were co-authored with overseas collaborators.

Number of publications

The table summarises the published output by our investigators and research groups that appeared with a 2007 publication date. We do not count papers that at the end of December 2007 were available only electronically in advance of print publication, or were in press, under review, submitted or in preparation. We provide two counts of publications, one (“Existing”) excluding publications by researchers associated with, but not technically part of, our Centre, and a second (“Extended”) that captures some, but certainly not all, of the related publications by our incoming CIs as listed in Appendix 1: Publications and Papers.

<table>
<thead>
<tr>
<th>Category</th>
<th>Existing</th>
<th>Extended</th>
</tr>
</thead>
<tbody>
<tr>
<td>B Book chapter</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>C1 Journal articles – articles in scholarly refereed journal</td>
<td>66</td>
<td>21</td>
</tr>
<tr>
<td>C2 Journal articles - other contribution to refereed journal</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>E1 Conference – full written paper – refereed proceedings</td>
<td>39</td>
<td>3</td>
</tr>
<tr>
<td>E2 Conference – full written paper – non-refereed proceedings</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>E4i Conference – edited volume of conference proceedings</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>E4 Conference – unpublished presentations</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>G Computer software</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>I Patents</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>– Other academic outputs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– demos</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>– posters</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>– reports</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>– postgraduate theses</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>TOTAL</td>
<td>166</td>
<td>30</td>
</tr>
</tbody>
</table>

Patents and trademarks

One patent was awarded during 2007 to Partner Investigator Isidore Rigoutsos. Previously we filed for registration of Visible Cell™ as an international trademark in Australia, New Zealand, USA, Europe and the United Kingdom; in 2007 we filed for registration in China. We own relevant Web domain names including www.visiblecell.org and www.visiblecell.com.

IMBcom Pty Ltd, a technology management corporation owned by The University of Queensland, conducted their annual three-day BioBusiness retreat in June 2007 and an introductory workshop in November 2007. The introductory workshop provides first-year PhD students with an overview of the concepts of bio-business, while the retreat for third-year students deals with issues in commercialisation and the diversity of career options. Several months prior to the retreat, students are assigned to groups and prepare, with the help of an IMBcom mentor, a business plan for presentation at the retreat. These two BioBusiness events together 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 students attending the three-day retreat in 2007 were:
- John Hawkins (University of Queensland)
- Mario Inostroza-Ponta (University of Newcastle)
- Shev MacNamara (University of Queensland)
Key Performance Indicators

**Invitations to address and participate in conferences**

**International conferences**

**Chief and Partner Investigators**
Dr Timothy Bailey
- Program Committee, Intelligent Systems and Molecular Biology (ISMB) 2007, Vienna, Austria (July)
- Program Committee, Neural Information Processing Systems Conference (NIPS) 2007, Whistler, Canada (December)
- Annual Symposium of the Association of Asian Societies for Bioinformatics (AASB/2007), Singapore (December)
- Program Committee, 2007 International Symposium on Computational Models for Life Sciences (CMLS'07), Gold Coast (December)

Prof. Kevin Burrage
- 32nd Annual Conference of South African Society for Numerical Mathematics and Applied Mathematics (SANUM 2007), Mateland, South Africa (April)
- International Conference on Computational Science (ICCS 2007), Beijing (May)
- International Conference on Scientific Computation and Differential Equations (SciCADE 07), St Malo, France (July)
- Sixth International Congress on Industrial and Applied Mathematics (ICIAM 07), Zürich, Switzerland (July)

A/Prof. Phoebe Chen
- Steering Committee Chair, Sixth Asia-Pacific Bioinformatics Conference (APBC2007), Hong Kong (January)
- Steering Committee Chair, 13th International Multimedia Modelling Conference (MMM2007) Singapore (January)
- Program Committee, 1st International Conference on Bioinformatics Research and Development (BIRD’07), Berlin, Germany (March)
- Area Chair, 11th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2007), Nanjing, China (May)
- Program Committee, 7th IEEE International Conference on Multimedia and Expo (ICME2007), Beijing China (July)

**Session Chair, 6th International Conference on Computer and Information Science (ICIS 2007), Melbourne (July)**

Prof. Michael Fellowes
- Computation and Logic in the Real World, Third Conference on Computability in Europe (CIE 2007), Siena, Italy (June)
- 13th Annual International Computing and Combinatorics Conference (COCOON 2007), Banff, Canada (July)
- 16th International Symposium of Fundamentals of Computation Theory (FCT 2007), Budapest, Hungary (August)
- Combinatorial Optimization and Applications, First International Conference, (COCOA 2007), Xi’an, China (August)

Dr Markus Hegland
- Sixth International Congress on Industrial and Applied Mathematics, Zürich, Switzerland (July)

**Session Chair, 2nd International Conference on Knowledge Science, Engineering and Management (KSEM 2007), Melbourne (November)**

Prof. Geoff McLachlan
- 56th Meeting of the International Statistical Institute (ISI), Lisbon, Portugal (August)
- Annual meeting of the Spanish Society of Statistics and Operational Research, Valladolid, Spain (September)
- The Sixth Biennial Scientific Meeting of the Classification and Data Analysis Group of the Italian Statistical Society (CLADAG SIS 2007), Macerata, Italy (September)
- Keynote, International Symposium on Computational Models for Life Sciences (CMLS) 2007, Gold Coast (December)

Prof. Bernard Palithope
- 8th Asia-Pacific Complex Systems Conference (Complex’07), Gold Coast (July)
- US-Australian Leadership Dialogue Meeting, Melbourne (August)

**Prof. Mark Ragan**
- Program Committee, Sixth Asia-Pacific Bioinformatics Conference (APBC2007), Hong Kong (January)
- First Bertinoro Systems Biology Meeting (BSB) 2007, Bertinoro, Italy (May)
- Plenary Session Co-Organiser and Speaker, Society for Molecular Biology and Evolution Conference (SMBE 2007), Halifax, Canada (June)
- 20th (final) Annual Meeting, Canadian Institute for Advanced Research (CIAR) Program in Evolutionary Biology, Halifax, Canada (June)
- FordFest International Symposium on the occasion of the retirement of Professor W. Ford Doolittle, Halifax, Canada (June)
- South Pacific User Services Conference (SPUSC) 2007, Gold Coast (November)
- Annual Symposium of the Association of Asian Societies for Bioinformatics (AASB/2007), Singapore (December)
- Program Committee, 18th International Conference on Genome Informatics, Singapore (December)
- Keynote, 2007 International Symposium on Computational Models for Life Sciences (CMLS’07), Gold Coast (December)

Dr Isidore Rigoutsos
- Keystone Symposium on microRNAs and siRNAs: Biological Functions and Mechanisms, Keystone CO (January)
- Keystone Symposium on RNAi for Target Validation and as a Therapeutic, Keystone CO (January)
Prof. Janet Wiles (Project 2.5)
- Keynote, 9th European Conference on Artificial Life, Lisbon, Portugal (September)
- Plenary, International Conference on Complex Systems (ICCS 2007), Boston, USA (October)

Prof. Xiaofang Zhou
- Program Committee, 13th International MultiMedia Modelling Conference (MMM 2007), Singapore (January)
- 18th Australasian Database Conference (ADC 2007), Ballarat (January)
- 23rd International Conference on Data Engineering (ICDE 2007), Istanbul (April)
- Program Committee, 12th International Conference on Database Systems for Advanced Applications (DASFAA 2007), Bangkok, Thailand (April)
- ACM SIGMOD International Conference on Management of Data (ACM SIGMOD 2007), Beijing, China (June)
- Program Committee, Joint Conference of the 9th Asia-Pacific Web Conference and the 8th International Conference on Web-Age Information Management (APWeb/WAIM 2007), Huang Shan, China (June)
- Program Committee, Joint Conference of the 9th Asia-Pacific Web Conference and the 8th International Conference on Web-Age Information Management (APWeb/WAIM 2007), Beijing, China (July)

Dr. Mikael Bodén
- Track Chair, 8th Asia-Pacific Complex Systems Conference (Complex’07), Gold Coast (July)
- Program Committee/Session Chair, Computational Models for 2007 International Symposium on Computational Models for Life Sciences (CMLS’07), Gold Coast (December)

Dr. Durgaprasad Bollina
- Alternative Splicing Special Interest Group meeting of the 15th Annual International Conference on Intelligent Systems for Molecular Biology & 6th European Conference on Computational Biology (ISMB/ECBB 2007), Vienna, Austria (July)
- 16th International Conference on Computing (CIC 2007), Mexico City (November)

Mr. Cheong Xin Chan
- Society for Molecular Biology and Evolution (SMBE 2007), Halifax, Canada (June)

Dr. Aaron Darling
- Annual New Zealand Phylogenetics Conference (Doom 07), Ruapehu, New Zealand (February)
- Session Chair, Evolution 2007, Christchurch (June)
- Program Committee, 5th Annual RECOMB Satellite Workshop on Comparative Genomics, San Diego, USA (September)

Dr. Nicholas Hamilton
- IBRO 2007 World Congress Satellite: The Secretory Vesicle Cycle and Novel Approaches to its Analysis, Brisbane (July)
- ACB-BINZ Bioinformatics Student Symposium, Brisbane (October)
- Program Committee, International Symposium on Computational Models for Life Sciences (CMLS’07), Gold Coast (December)

Mr. Mohammad T. Islam
- International Conference on Information and Communication Technology (ICICT 2007), Dhaka (March)

Dr. Karin Kassahn
- Evolution 2007, Christchurch, NZ (June)

Dr. Lesheng Kong
- 6th International Conference on Bioinformatics (InCoB 2007), Hong Kong (August)

Mr. Shev MacNamara
- Sixth International Congress on Industrial and Applied Mathematics (ICIAM 07), Zürich (July)

Mr. Shivashankar Nagaraj
- 15th Annual International Conference on Intelligent Systems for Molecular Biology & 6th European Conference on Computational Biology (ISMB/ECBB 2007), Vienna (July)

Dr. Joo Chuan Tong
- 6th International Conference on Bioinformatics (InCoB 2007), Hong Kong (August)

Dr. James Watson
- 8th Asia Pacific Complex Systems Conference (Complex’07), Gold Coast (July)
National conferences

Chief Investigators
Prof. Kevin Burrage
- Scientific Programme Committee, Bioinformatics Australia, Brisbane (October)

A/Prof. Phoebe Chen
- Bioinformatics Australia 2007 Conference, Brisbane (October)

Prof. Geoff McLachlan
- 2007 Winter School in Mathematical and Computational Biology, Brisbane (June)

Prof. Bernard Pailthorpe
- APAC 2007 Conference, Perth (October)
- eResearch Australasia Conference, Brisbane (June)
- Legal Frameworks for eResearch Conference, Gold Coast (July)
- SPIE Nanotechnology Conference, Canberra (December)

Prof. Mark Ragan
- BCCG (Bioinformatics Capability Consultative Group) Committee Meeting, Sydney 16 March 2007
- Sydney Bioinformatics ComBio Satellite Meeting, Sydney (September)
- Scientific Programme Committee, Bioinformatics Australia, Brisbane (October)
- BioInfoSummer, Canberra (December)

Prof. Shoba Ranganathan
- Bioinformatics Australia 2007 Conference, Brisbane (October)

Dr Rohan Teasdale
- AusBiotech National Conference, Brisbane (October)
- Chair, Organising Committee, Bioinformatics Australia 2007 Conference, Brisbane (October)
- Chair, Scientific Programme Committee, Bioinformatics Australia 2007 Conference (October)

Prof. Janet Wiles (Project 2.5)
- Program Committee Co-Chair, 8th Asia-Pacific Complex Systems Conference (Complex’07), Gold Coast (July)

Postdoctoral fellows, Research Officers and students
Dr Michael Bodén
- Session Chair, Bioinformatics Australia, Brisbane (October)
- Program Committee, Twentieth Australian Joint Conference on Artificial Intelligence (AI 2007), Gold Coast (December)
- Mr Cheong Xin Chan, Dr Aaron Darling, Dr Melissa Davis, Dr Nicholas Hamilton, Dr Karin Kassahn, Dr Shaob Sehgal, Mr Chang Jin Shin, Mr David Wood, Dr Simon Wong, & Dr Zheng Yuan
- Bioinformatics Australia 2007 Conference, Brisbane (October)

Dr Nicholas Hamilton
- Chair of Organisation Committee and Symposium chair (“Biological Imaging”), 2007 Winter School in Mathematics and Computational Biology, Brisbane (June)
- Dr Karin Kassahn
- BioInfoSummer 2007, Canberra (December)

Mr Tom Whittington
- BioInfoSummer 2007, Canberra (December)

Dr Ian Wood
- Session Chair, 51st Annual Meeting of the Australian Mathematical Society, Melbourne (September)

Invitations to visit leading international laboratories, major research groups and centres

Chief Investigators
Dr Timothy Bailey
- William Noble Lab, University of Washington, Seattle (November)
- Limsoon Wong Lab, National University of Singapore, Singapore (December)

Prof. Kevin Burrage
- African Institute for Mathematical Sciences, Cape Town, South Africa (March)
- Chinese Academy of Sciences, Beijing (May)
- Bioinformatics Centre, Barcelona (July)

A/Prof. Phoebe Chen
- Department of Computer Science, National University of Singapore, Singapore (January)
- Institute for Computational Biology, Chinese Academy of Sciences, Shanghai (May)
- Department of Computer Science & Engineering, Fudan University, Shanghai (May)
- Auckland Bioengineering Institute, University of Auckland (November)
- Bioinformatics Institute, University of Auckland (November)

Prof. Michael Fellows
- Institute of Advanced Studies, Durham University, UK (January–March)
- Dr Iris van Rooij, Technical University of Eindhoven, Netherlands (March)
- Parameterized and Exact Computation Group, University of Bergen, Norway (April–June)
- Prof. Hans Bodlaender, University of Utrecht, Netherlands (July)
- Prof. Hanning Fernau, Department of Computer Science, University of Trier, Germany (July)
- University of Jena, Germany (August)
- Prof. Martin Grohe, Humboldt University, Berlin (September)
- Dr Christian Knauer, Free University in Berlin (September)
- Prof. Detlef Seese, University Karlsruhe, Germany (September)
- Prof. Juraj Hromkovic, ETH, Zürich (October)

Dr Markus Hegland
- Institute for Numerical Simulation, University of Bonn, Germany (August–October)
- RICAM, Austrian Academy of Science, University of Linz, Austria (November)
- Technische Universität Berlin, Germany (September)

Prof. Geoff McLachlan
- Department of Statistics, University of Bologna, Italy (September)
- Symposium on mixture modelling with special interest to applications in educational measurement and bioinformatics, Research Group of Quantitative Psychology and Individual Differences, University of Leuven, Belgium (November)

Dr Brad Marsh
- Faculty of Life Sciences, Brandeis University, Boston (December)
Prof. Bernard Pailthorpe
• San Diego Supercomputer Center (May)
• University of California San Diego (May)
• California Institute for Telecommunications and Information Technology (CalIt2) (May)
• INSEET, Laboratoire Souterrain Bas Bruit de Rustrel-Pays d’Apt, France (November)

Prof. Mark Ragan
• Biotechnology Research Institute, National Research Council Canada, Montréal (June)
• Department of Biochemistry, McGill University, Montréal (June)
• Institute for Marine Bioscience, National Research Council Canada, Halifax (June)

Prof. Shoba Ranganathan
• Knowledge Discovery Group, Institute for Infocomm Research, Singapore (March)
• MD Anderson Cancer Center, Houston, USA (June)
• School of Biological Sciences, Nanyang Technological University, Singapore (September)

Dr. Rohan Teasdale
• Fantom4.5, RIKEN Genome Research Centre, Tokyo (May)

Prof. Janet Wiles (Program 2.5)
• NSF-funded Temporal Dynamics of Learning Centre, University of California, San Diego (April–June)

Postdoctoral fellows, Research Officers and students
Dr. Aaron E. Darling
• Jonathan Eisen laboratory, University of California, Davis (May–June)
• Bioinformatics Institute, University of Auckland, New Zealand (June–July)

Ms Chin Foon Khoo
• NICTA, Melbourne University, Melbourne, September

Dr James Watson
• Biosystems Research Group, School of Computing, University of Leeds, UK (September)
• Centre for Research in Language, University of California, San Diego (October)

Commentaries about the Centre’s achievements
Dr. Sean Grimmond was featured in The Weekend Australian (“Two DNA sequencers a step ahead”, 15–16 December 2007 pages 13 & 14) in regard of his introduction of the Applied Biosystems next-generation DNA sequencing technology into Australia.

The Visible Cell™ project featured in two presentations to the senior executive group from the International Federation for Information Processing as part of the bid by Brisbane to host the IFIP World Congress of IT in 2010. Brisbane’s bid was successful against competition that included Bangalore, Beijing and Lisbon.

The Visible Cell™ project and personnel were also featured in the video highlighting the University of Queensland’s Research Week 2007 (12–17 September). The video introduced the UQ Foundation Research Excellence Awards (20 September 2007) at which Professor Margaret Sheil, CEO of Australian Research Council, was the guest speaker.

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 mathematics, or bioscience and ICT. In this way we 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 multidisciplinary teams.

Of our 12 projects and sub-projects active during 2007, 10 (83%) included personnel with advanced degrees in different disciplines; 18 of our 21 PDSs (86%), 15 of our 35 PhD students (43%), 2 Masters students and 2 international interns worked in these 12 projects. Of our 50 papers and posters during 2007 with at least two Centre investigators or students as co-authors, 29 (58%) of those co-authors hold advanced degrees in different disciplines. One-third (7 of 21) of our 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.
2. Research Training and professional education

Recruitment
In 2007, ACB recruited:

- Dr Mikael Bodén, Senior Research Fellow, 3-year secondment from UQ School of ITEE for the primary purpose of developing and delivering advanced coursework and projects in bioinformatics. Mikael will continue to collaborate in Project 2.2 (Computational discovery of gene-regulatory motifs and motif clusters).
- Dr Pamela Burrage and Dr Krzysztof Kurowski, Research Officers, in grid services for complex systems simulations to support computational modelling in Projects 2.4 (Mathematical modelling of gene regulatory networks), 2.5 (Complex systems modelling of regulatory and protein-protein interaction networks), and 3.2.2 (Modelling and visualisation of biological processes in complex spatial environments).
- Mr Cheong Xin Chan, Research Officer, into Project 2.6 (Ortholog mapping of prokaryotic and eukaryotic genomes).
- Dr Melissa Davis, Research Officer, into Project 3.1.1 (Data environment for the Visible Cell™ project).
- Mr Lloyd Flack, Research Assistant, into Project 2.3 (New statistical approaches in bioinformatics).
- Dr Timothy McComb, Application Developer, into Project 3.2.1 (Concept and engineering of the Visible Cell™).
- Dr Mohamad Shoail Sehgal, Research Officer, into Project 3.1.2 (Integrating database technologies and visual analysis for genome data).
- Dr Needlima Pottekkat Sidharthan, Research Officer, into Project 2 (Data and process management for the Visible Cell™) of the extended Centre’s Program 1 (The Visible Cell™, linking genome and phenotype).
- Dr Ian Wood, Research Officer, into Project 2.3 (New statistical approaches in bioinformatics).

Professional/technical training and advanced career development

Seminar/Symposium hosted and/or sponsored by ACB
In 2007, ACB hosted and sponsored:

- CILR/ACB Joint Seminar (23 April 2007)
- 2007 Winter School in Mathematical & Computational Biology (25–29 June 2007)
- Bioinformatics Australia 2007 meeting (23–24 October 2007)
- 2007 ACB-BINZ Bioinformatics Student Symposium (25–26 October 2007)
- 2007 ACB Annual All-Hands Meeting (28–27 November 2007)

CILR/ACB Joint Seminar (23 April 2007)
The ARC Centre of Excellence for Integrative Legume Research (CILR) and the ARC Centre of Excellence in Bioinformatics jointly hosted a seminar on 23 April 2007. Dr Georg Weiller (Australian National University) presented a lecture on “Bioinformatics approaches to study gene expression and interaction in the CILR”. Both Centres’ postdoctoral fellows and students attended.

2007 Winter School in Mathematical and Computational Biology (25–29 June 2007)
The 2007 Winter School in Mathematical and Computational Biology featured 27 speakers (5 international, 21 national and 1 commercial). We welcomed 190 participants from 35 distinct institutions, mostly from Australia and New Zealand. Approximately 68% were researchers or postgraduate students, and 32% advanced undergraduates. One-third identified a biological background, with most others from mathematics, statistics, computer science or IT. About 80% represented Queensland universities. Again, we received highly positive written feedback from the participants about the topics and quality of presentations.

We thank Queensland Cyber Infrastructure Foundation (QCIIF), The MathWorks Australia, SGI and the Institute for Molecular Bioscience for their generous sponsorship of our Winter School.
Day 3  
Prediction and modelling of protein structure and dynamics  
Prof. Alan Mark  
ARC Federation Fellow  
Centre in Computational Molecular Sciences  
School of Molecular and Microbial Sciences, and Institute for Molecular Bioscience  
The University of Queensland  
Dr Thomas Huber  
School of Molecular and Microbial Sciences  
The University of Queensland  
Prof. Bernard Pailthorpe, CEO  
Queensland Cyber Infrastructure Foundation  
ARC Centre in Bioinformatics  
The University of Queensland  
Prof. Phil Bourne  
Co-Director, Protein Data Bank (PDB)  
San Diego Supercomputer Center, and  
Department of Pharmacology  
University of California, San Diego, USA  

Day 4  
Statistical analysis of gene expression  
Prof. Sue Wilson  
Co-Director, Centre for  
Bioinformatics Science  
The Australian National University  
Dr Gordon Smyth  
Senior Research Scientist  
Bioinformatics Division  
Walter and Eliza Institute for  
Medical Research  
Prof. Geoff McLachlan  
ARC Centre in Bioinformatics, and  
Department of Mathematics  
The University of Queensland  
Dr Antonio Reverter  
Principal Research Scientist  
Bioinformatics Group  
CSIRO Livestock Industries  

Dr Harri Kiiveri  
Statistical Bioinformatics – Health  
CSIRO Mathematical and Information  
Sciences, Floreat, Western Australia  
Dr Ian Wood  
Research Fellow in Mathematical Sciences  
Queensland University of Technology  
Prof. Phil Bourne (plenary lecture)  
Co-Director, Protein Data Bank (PDB)  
San Diego Supercomputer Center, and  
Department of Pharmacology  
University of California, San Diego, USA  

Day 5  
Computational neuroscience  
A/Prof. Geoff Goodhill  
Queensland Brain Institute  
The University of Queensland  
Dr Tatyana Sharpee  
Salk Institute for Biological Studies  
San Diego, USA  
Prof. Mandym Srinivasan  
ARC Federation Fellow  
Queensland Brain Institute  
The University of Queensland  
Dr Anthony Bell  
Redwood Center for Theoretical  
Neuroscience  
The University of California, USA  
A/Prof. Michael Breakspear  
School of Psychiatry  
The Black Dog Institute  
University of New South Wales

2007 ACB–BINZ Bioinformatics Student Symposium (25–26 October 2007)  
The second joint Student Symposium in  
Bioinformatics with Bioinformatics Institute New Zealand (BINZ) was held in Brisbane  
(25–26 October 2007). This year, ACB  
sponsored the venue, local arrangements  
and accommodation for the New Zealand  
participants; 40 students and staff from BINZ,  
ANU, Deakin University, The University of  
Newcastle and The University of Queensland  
participated. The student symposium  
program included short presentations from  
students, and longer presentations from  
established researchers.

Chief Investigators and associates:  
Dr Timothy Bailey  
Dr Mikael Bodén  
Prof. Kevin Burrage  
Dr Melissa Davis  
Dr Nicholas Hamilton  
Prof. Geoff McLachlan  
Dr Brad Marsh  
Prof. John Mattick  
Prof. Mark Ragan  
Dr Rohan Teasdale

ACB postgraduate students:  
Farah Abdullah (UQ)  
Isye Arieshanti (UQ)  
Alhadi Bustamam (UQ)  
Chong Xin Chan (UQ)  
JooYoung Choi (UQ)  
Mario Inostroza-Ponta (Newcastle)  
Chin Foon Khoo (ANU)  
Shev MacNamara (UQ)  
Scott Mann (Deakin)  
Andrew Noske (UQ)  
Chang Jin Shin (UQ)  
Cas Simons (UQ)  
Josephine Sprenger (UQ)  
Peter van der Heide (UQ)  
Tom Whittington (UQ)

2007 ACB–BINZ Bioinformatics Student Symposium (25–26 October 2007)
ACB 2007 Annual All-Hands Meeting (26–27 November 2007)

ACB Chief Investigators, staff and students participated in the Annual All-Hands Meeting. On the first day, the existing Centre’s Chief Investigators and postdocs presented a brief recapitulation of major outcomes of their respective research programs. On Day 2, Chief Investigators in the extended Centre of Excellence met to plan the upcoming research program for 2008 and beyond, and to identify specific steps and timelines in the transition.

Other research and training and professional education

Apart from the Winter School and Student Symposium, our CIs and researchers participated in:

Dr Nicholas Hamilton, Dr Markus Kerr, Dr Suzanne Norwood, Dr Rohan Teasdale & Mr Jack Wang

• QIMR Scientific Imaging Workshop (May)

Postgraduate and undergraduate courses in the Centre’s area

Our investigators and researchers participated in the following for-credit courses and programs in 2007:

A/Prof. Phoebe Chen
  • SIT781 Introduction to Bioinformatics (unit chair and lecturer)
  • SIT772 Database and Information Retrieval (lecturer)
  • SIT790 MIT Professional (Research) (unit chair)

Dr Sean Grimmond
  • BIOL3004 Genomics and Bioinformatics (lecturer)
  • BIOT7009 Emerging Technologies (lecturer)

Dr Karin Kassahn
  • BIOL3012, Introduction to Research (project supervisor)

Dr Brad Marsh
  • BIOL3006 Molecular Cell Biology (lecturer)
  • BIOC3004 Structural Biology: Maromolecular Structure (lecturer)

Prof. Bernard Pailthorpe
  • MATH3201 Scientific Computing – Advanced Techniques and Applications (organiser and lecturer)
  • MATH4201 Advanced Modelling and Scientific Visualisation (lecturer)

Prof. Mark Ragan
  • BIOL3004 Genomics and Bioinformatics (introductory lecture)

Prof. Shoba Ranganathan
  • CBMS880 (Macquarie University), Special Topics in Biotechnology (lecturer)
  • CBMS301 (Macquarie University), Technology Miniproject (convenor)
  • BS413 (Nanyang Technological University, Singapore), Immunomics (special lecturer)

Dr Rohan Teasdale
  • BIOL2004 Genomics and Bioinformatics (lecturer)
  • BIOC6000 Advanced Genomics & Bioinformatics (project supervisor)

Dr Ian Wood
  • COMP3702/COMP7702 Artificial Intelligence (lecturer and course coordinator)

Senior high school student research scheme

Commitments have been engaged in the following outreach programs to secondary school science students:

CSIRO Student Research Scheme

Dr Melissa Davis and Dr Simon Wong supervised four secondary-school students (Years 11–12) under auspices of the CSIRO Student Research Scheme. The two-week project introduced students to the scientific research environment and to bioinformatics as an interdisciplinary research area.

Queensland Academy Biotechnology Research Experience Program (QABREP)

Prof. Mark Ragan participated in planning for ACB to host approximately 12 students from the new Queensland Academy of Science, Mathematics and Technology in the QASMT’s inaugural Biotechnology Research Experience Program (QABREP) commencing in 2008. Two students at a time will be progressively rotated through different laboratories in the Faculty of Biological & Chemical Sciences, the Australian Institute of Biomaterials and Nanotechnology, and the ACB node in the Institute for Molecular Bioscience, UQ, spending one afternoon per week per laboratory over a 6- to 8-week period.

Scientists in Schools Partnership

Dr Brad Marsh is the Centre’s point of contact for the Scientists in Schools Partnership program. This program, funded by the Australian Government Department of Education, Science and Training, is led by Dr Jim Peacock, Australia’s Chief Scientist, and managed by CSIRO. The program identifies a school that matches closely with the scientist’s preference for location, school level, interests and preferred subjects. Two of our new CI (Dr Brad Marsh and Prof. Shoba Ranganathan) and postdoctoral fellows Dr Karin Kassahn and Dr Melissa Davis have registered to participate in the program.
3. International, national and regional links and networks

Visitors

International visitors
- Mr Jeff Adie, Senior Principal Engineer, SGI (Singapore)
- Dr Anthony J. Bell, Computational Neurobiology Lab, The Salk Institute, USA
- Dr Manuel Barrio, Valladolid University, Spain
- Prof. Werner Dubitzky, Head, Bioinformatics Program, University of Ulster
- Dr Murray Grigor, University of Auckland
- Dr Michael Hucka, Co-Director, Biological Network Modelling Center, California Institute of Technology
- Prof. Peter Hunter, Director, Bioengineering Institute, University of Auckland
- Dr Philip Kim, Department of Medical Biophysics & Biochemistry, Yale University
- Dr Krzysztof Kurowski, Poznan Supercomputing and Networking Center, Polish Academy of Sciences
- Prof. David Lambert FRSNZ, Massey University
- Prof. Sang Yup Lee, Director, Bioinformatics Research Centre, KAIST (Korea)
- Dr David Levin, Tel Aviv University
- Prof. Gregory McRae, Department of Chemical Engineering, Massachusetts Institute of Technology
- A/Prof. Osamu Maruyama, Faculty of Mathematics and Graduate School of Systems Life Sciences, Kyushu University, Japan
- Prof. Yoshinobu Mineyuki, Department of Life Science, University of Hyogo, Japan
- Prof. Robert F. Murphy, Computational Cell Biology program, Carnegie Mellon University
- Mr Dan Nicolau Jr (PhD student), Oxford University
- Dr Sergey Piskarev, Science Research Computer Center, University of Moscow
- Prof. Hoang Quang, Hue University, Vietnam
- Prof. Greg Rouse, Scripps Oceanographic Institute, San Diego
- Prof. Brian Storrie, University of Arkansas for Medical Sciences, Little Rock
- Dr Larry Smarr, Director, California Institute for Telecommunications and Information Technology (CalIT2), University of California San Diego
- Prof. Ah Chung Tsoi, Monash University/Baptist University Hong Kong
- Mr Todd Treangen, Universitat Politècnica de Catalunya, Barcelona
- Prof. David Levin, Tel Aviv University
- Prof. Gregory McRae, Department of Chemical Engineering, Massachusetts Institute of Technology
- A/Prof. Osamu Maruyama, Faculty of Mathematics and Graduate School of Systems Life Sciences, Kyushu University, Japan
- Prof. Yoshinobu Mineyuki, Department of Life Science, University of Hyogo, Japan
- Prof. Robert F. Murphy, Computational Cell Biology program, Carnegie Mellon University
- Mr Dan Nicolau Jr (PhD student), Oxford University
- Dr Sergey Piskarev, Science Research Computer Center, University of Moscow
- Prof. Hoang Quang, Hue University, Vietnam
- Prof. Greg Rouse, Scripps Oceanographic Institute, San Diego
- Prof. Brian Storrie, University of Arkansas for Medical Sciences, Little Rock
- Dr Larry Smarr, Director, California Institute for Telecommunications and Information Technology (CalIT2), University of California San Diego
- Prof. Ah Chung Tsoi, Monash University/Baptist University Hong Kong
- Mr Todd Treangen, Universitat Politècnica de Catalunya, Barcelona
- Prof. David Levin, Tel Aviv University
- Prof. Gregory McRae, Department of Chemical Engineering, Massachusetts Institute of Technology
- A/Prof. Osamu Maruyama, Faculty of Mathematics and Graduate School of Systems Life Sciences, Kyushu University, Japan
- Prof. Yoshinobu Mineyuki, Department of Life Science, University of Hyogo, Japan
- Prof. Robert F. Murphy, Computational Cell Biology program, Carnegie Mellon University
- Mr Dan Nicolau Jr (PhD student), Oxford University
- Dr Sergey Piskarev, Science Research Computer Center, University of Moscow
- Prof. Hoang Quang, Hue University, Vietnam
- Prof. Greg Rouse, Scripps Oceanographic Institute, San Diego
- Prof. Brian Storrie, University of Arkansas for Medical Sciences, Little Rock
- Dr Larry Smarr, Director, California Institute for Telecommunications and Information Technology (CalIT2), University of California San Diego
- Prof. Ah Chung Tsoi, Monash University/Baptist University Hong Kong
- Mr Todd Treangen, Universitat Politècnica de Catalunya, Barcelona

National visitors to one or more ACB nodes
- Prof. Alan Bell, Chief, CSIRO Livestock Industries
- Ms Leanne Bischof, CSIRO Mathematical and Information Sciences
- Mr Nick Conomo, SGI
- Mr Brad Guthbert, Information Industries Bureau, Queensland Government
- Mr Brian Demnar, Information Industries Bureau, Queensland Government
- Prof. Jenny Graves, Research School of Biological Sciences, ANU
- Dr Anna Guppy, Queensland Academy of Science, Mathematics & Technology
- Ms Mary Lloyd, Information Industries Bureau, Queensland Government
- Mr Stewart Macintyre, Information Industries Bureau, Queensland Government
- Ms Sharmila Mercer, Mercorp Consulting
- Dr David Mitchell, CSIRO Mathematical & Information Sciences
- A/Prof. Tuan Pham, Head of Bioinformatics Program, James Cook University
- Ms Angela Rehrl, Perkins Resources
- Prof. Nadia Rosenthal, Director, Australian Regenerative Medicine Institute
- Mr Robert Wall, Business Manager for Asia Pacific, Mitrionics Ltd.
- Dr Georg Weiller, Australian National University
- Dr Bill Wilson, CSIRO Mathematical & Information Sciences

Sabbatical visitors
- Mr Hong Liang Hiew, Centre for Comparative Genomics, Murdoch University
- Prof. Wendy Wen-Rong Jiang, Shanghai Second Polytechnics University, China
- A/Prof. Mitsuki Nakasumi, Komazawa University, Japan
- A/Prof. Hyu Chan Park, Korea Maritime University, Korea

From left: Dr Nicholas Hamilton, Prof. Kevin Burrage, Prof. Mark Ragan, Dr Michael Hucka and Prof. John Mattick during Dr Hucka’s visit 25–26 October 2007.
Collaborative national and international workshops and exchanges

In addition to the 2007 Winter School and Student Symposium, our CIs, PDFs, PhDs presented at or participated in workshops as follows:

Chief Investigators
A/Prof. Phoebe Chen
- Program Committee, 2007 Workshop on Data Mining for Biomedical Applications (BioDM2007), Nanjing (May)
- 7th International Workshop on Data Mining in Bioinformatics (BIODDD07), San Jose (August)
- Program Committee, 4th Integrative Bioinformatics Workshop (IB2007), Belgium (September)
- Program Committee, 2nd Workshop on Pattern Recognition in Bioinformatics (PRIB’07), Singapore (October)
- Conference Co-Chair, First International Workshop on Conceptual Modelling for Life Sciences Applications (CMLSA2007), Auckland (November)

Prof. Michael Fellows
- Third Bellairs Workshop on Parameterized Algorithms, Holetown, Barbados (February)
- London Algorithms Workshop (LAW 2007), London (February)
- Keynote, Fourth Dagstuhl Workshop on Parameterized Complexity, Dagstuhl Castle, Germany (July)
- Second Workshop on Algorithms and Complexity (ACiD 2007), Durham (September)

Dr Brad Marsh
- ARC Centre of Excellence for Coherent X-ray Science (CXS) 2nd Annual Workshop, Bio21 Institute, Melbourne (April)
- 8th OIE/WAVLD Seminar on Biotechnology: Applications of Biotechnology to the Diagnosis and Pathology of Animal Diseases, Melbourne (November)
- 47th American Society for Cell Biology Annual Meeting, Mini-symposium on High-Tech Cell Biology, Washington (December)

Prof. Bernard Pailthorpe
- Access Grid Workshop, Chicago (May)

Prof. Mark Ragan
- QosCosGRID Technical meeting, French Ministry of Research, Paris (February–March)
- NBCF (National Breast Cancer Foundation) Workshop, Sydney (April)
- Bioinformatics Australia Consultative Workshop (September)
- CISRO Transformational Bioinformatics meeting (November)
- SRS and Cowrie Web Services Training (November)
- UQ Grant-Writing Seminar for Biological Sciences and Biotechnology (December)

Prof. Shoba Ranganathan
- Grid Computing and Bioinformatics workshop, Auckland (May)

Prof. Janet Wiles
- Dynamics of Development workshop, Portugal (September)

Prof. Xiaofang Zhou
- The Second IEEE International Workshop on Multimedia Databases and Data Management (MDDM’07), Istanbul (April)
- Program Committee, 5th International Workshop on Biological Data Management (BIoDM’07), Regensburg, Germany (September)

Postdoctoral fellows, Research Officers and students

Dr Mikael Bodén
- 2nd IAPR Workshop on Pattern Recognition in Bioinformatics, Singapore (October)

Dr Aaron Darling
- St John’s Workshop on Advances in Bacterial Genomics, Oxford, UK (September)

Dr Melissa Davis and Dr Shoaib Sehgal
- 2nd eMBL (electronic International Molecular Biology Laboratory) Virtual Workshop for Systems Biology via AccessGrid (December 2007)

Dr Nicholas Hamilton
- Program Committee, Workshop on Intelligent Systems in Bioinformatics, Hobart (December)

Dr Jim Hanan
- Co-Chair, 5th International Workshop on Functional Structural Plant Models, Napier (November)

Mr Peter Shaw
- 33rd International Workshop on Graph-Theoretic Concepts in Computer Science, Dornburg (June)

Dr James Watson
- The 7th International Workshop on Information Processing in Cells and Tissues (IPCAT), Oxford (August)
- CompuSteer Workshop, The University of Hull, UK (September)

Mr Tom Whittington
- GeneSpring Workshop, Brisbane (July)
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 institutions as follows. Refer also to page 46, Invitations to visit leading international laboratories.

National collaborating institutions:

Australian Centre for Plant Functional Genomics
Australian Institute of Marine Sciences
Australian Regenerative Medicine Institute, Monash University
ARC Centre for Complex Systems
ARC Centre of Excellence for Integrative Legume Research
ARC Special Research Centre in Functional and Applied Genomics
Central Queensland University
Charles Sturt University
CSIRO Mathematical and Information Sciences
CRC Australian Weed Management, Department of Agriculture, WA
Curtin University of Technology
eHealth Research Centre, CSIRO
Energy Edge Pty Ltd
Griffith University
James Cook University
Macquarie University
Monash University
Murdoch University
National ICT Australia (NICTA)
Nielson Media Research Australia
Peter Macallum Cancer Centre
Prince Henry’s Institute, Adelaide
Queensland Cyber Infrastructure Foundation
Queensland Centre of Medical Research
Queensland University of Technology
Suncorp Metway
The Mechanics Institute, B’d’s Farm, NSW
The Walter and Eliza Hall Institute of Medical Research
University of Adelaide
University of New South Wales
University of Sydney
University of Melbourne
University of Technology, Sydney
University of Western Australia
Westmead Hospital, Sydney

International collaborating institutions:

Bioengineering Institute, University of Auckland
Bioinformatics Centre, Barcelona
Biotechnology Research Institute, National Research Council Canada
California Institute of Technology, USA
California Institute for Telecommunications and Information Technology (CalIT2)
Carleton University, Canada
Carson-Newman College, USA
Chinese Academy of Sciences, Beijing
Chinese University of Hong Kong
Chonnam National University, South Korea
Chulalongkorn University, Thailand
City University of Hong Kong
Dalhousie University, Canada
Dana Faber Cancer Institute, USA
Danish Technical University, Denmark
Dionne Complex Systems, UK
Durham University, UK
ETH, Zürich
Free University, Berlin
Fudan University, China
Genome Institute of Singapore
Harbin Institute of Technology, China
Harvard University, USA
Humboldt University, Germany
IBM T.J. Watson Research Center, USA
Indian Institute of Technology, India
Institute for Infocomm Research, Singapore
Institute of Advanced Studies, Durham University
Institute for Infocomm Research, Singapore
Institute of Mathematical Sciences, India
Johns Hopkins University School of Medicine, USA
KAIST (Korea Advanced Institute of Science & Technology), Korea
Laboratory for Molecular Biology, Medical Research Council, UK
Lebanese American University, Lebanon
Lund University, Sweden
Massachusetts Institute of Technology, USA
Massey University, New Zealand
McGill University, Montréal
MD Anderson Cancer Center, Houston, USA
Nanyang Technological University, Singapore
National University of Singapore
Ohio University, USA
Parahyangan Catholic University, Indonesia
Pitzer Research Technology Center, USA
Purdue University, USA
Rensselaer Polytechnic Institute, USA
RICAM, Austrian Academy of Science, Austria
RIKEN Genome Sciences Center, Japan
RIKEN Wako Institute, Japan
RIKEN Yokohama Institute, Japan
Sandia National Labs, USA
South China University of Technology, China
SRA International, Inc., USA
State University of New York, USA
University of Stellenbosch, South Africa
Technical University of Eindhoven, Netherlands
Technische Universität Berlin, Germany
Technische Universität Clausthal, Germany
Tel Aviv University, Israel
The Natural History Museum, UK
The Open University, Milton Keynes, UK
The University of Hong Kong
The University of Leeds, UK
The University of Newcastle upon Tyne, UK
Trinity College Dublin, Ireland
Tsinghua University, China
Università degli Studi di Perugia, Italy
Universitat Politècnica de Catalunya, Spain
Universität Bonn, Germany
Universität Jena, Germany
Universität Karlsruhe, Germany
Université de Rennes, France
Université de Nantes, France
Université Paris-Sud, France
University of Auckland, New Zealand
University of Bergen, Norway
University of Birmingham, UK
University of Bologna, Italy
University of California, Berkeley, USA
University of California, San Diego, USA
University of Calgary, Canada
University of Canterbury, New Zealand
University of Chester, UK
University College Cork, Ireland
University of Denver, USA
University of Georgia, USA
University of Hafia, Israel
University of Hong Kong, China
University of Iowa, USA
University of Leuven, Belgium
University of Macerata, Italy
University of Oxford, UK
University of Puerto Rico-Medical Science, Puerto Rico
University of South Carolina, USA
University of Southern California, USA
University of Trier, Germany
University of Tokyo, Japan
University of Tennessee, USA
University of Texas, USA
University of Ulster, UK
University of Urbino, Italy
University of Utrecht, Netherlands
University of Vienna, Austria
University of Washington, USA
University of Waterloo, Canada
University of Wisconsin, USA
Uppsala University, Sweden
Utrecht University, Netherlands
Xiamen University, China
Zhuhai State-owned Assets Supervision and Management Commission, China

Key Performance Indicators

2007 ANNUAL REPORT
ARC CENTRE OF EXCELLENCE IN BIOINFORMATICS

53
Membership on national and international professional committees

In addition to participation on organising or program committees for specific conferences or workshops, ACB Chief Investigators were actively engaged in 2007 with executive, management and professional committees, advisory and review committees, and similar continuing bodies (excludes journal editorial boards and learned societies):

A/Prof. Phoebe Chen
• Chair, Asia-Pacific Bioinformatics Conference Steering Committee
• Chair, International Conference of Multimedia Modelling Steering Committee
• Member, Steering Committee, Bioinformatics Australia

Prof. Michael Fellows
• Member, Steering Committee, International Workshop on Parameterized and Exact Computation

Dr Brad Marsh
• Chair, Communications & Program Awareness Sub-committee, Network for Pancreatic Organ Donors with Diabetes (Juvenile Diabetes Research Foundation, USA)

Prof. Bernard Pailthorpe
• Chair, NCRIS ARCS Executive Committee

Prof. Mark Ragan
• Chair, Executive group, Association of Asian Societies of Bioinformatics
• Vice-President, Bioinformatics Australia
• Chair, Management Committee, Queensland Facility for Advanced Bioinformatics
• Member, Board of Management, ARCHER (Australian Research Enabling Environment)
• Member, Advisory Board, Quasi-Opportunistic Supercomputing for Complex Systems in Grid Environments project, European FP6
• Member, Merit Allocation Committee, APAC National Supercomputing Facility
• External Examiner, University of Malaya Masters of Biotechnology Programme

Prof. Shoba Ranganathan
• President, Asia-Pacific Bioinformatics Network
• Chair, International Conference in Bioinformatics Executive Committee
• Chair, S* Life Science Informatics Alliance
• Member, Steering Committee, Bioinformatics Australia
• Member, Management Committee, Macquarie University Biotechnology Research Institute
• Member of Council, International Immunomics Society

Dr Rohan Teasdale
• Member, Steering Committee, International Conference on Bioinformatics
• Member, Steering Committee, Bioinformatics Australia

Prof. Xiaofang Zhou
• Executive Committee Member, Web Information Systems Engineering Society (WISE Society)
• Expert Consultant, Mid and Long Term Strategic Planning for Science and Technology Development, Ministry of Science and Technology of the People’s Republic of China
• Member, Technical Advisory Board, International Conferences on Web Services (ICWS)
• Member, Advisory Committee, Adaptive Service Agreement and Process Management in Services Grids Project, an EU-DEST project hosted by Swinburne University

Research projects with international partners

Listed below are research projects with international partners that are underway in 2007:

Dr Timothy Bailey
• A/Prof. William Noble, Department of Genome Sciences, University of Washington (USA)
MEME suite of motif-based sequence analysis tools

Prof. Kevin Burrage
• Dr Christopher Bouton & Dr Victor Farutin, Pfizer Research Technology Center (Cambridge)
BioManta: modelling and analysis of biological network activity.
• Prof. John Hancock (UQ) & Dr Dan Nicolau Jr, Oxford University (U.K)
Stochastic simulations for modelling anomalous diffusion and exploring the role of lipid rafts in the plasma membrane.
• Dr André Leier & Dr Tatiana Marquez Lago, ETH (Zürich)
Dynamics of the Her1 and Her 7 complex in somite formation in zebrafish.
• Dr Grant Lythe, University of Leeds (UK)
New numerical methods for second-order stochastic equations.
• Dr Tatiana Marquez Lago, ETH (Zürich)
Multiscale algorithms for the spatial modelling of discrete chemical kinetics in a cell.
• Manuel Barrio Solórzano, Departamento de Informática, Universidad de Valladolid, Valladolid, Spain
Dynamics of the Hes1 molecular clock in mouse.
• Dr Jiangning Song, Kyoto University (Japan)
Predicting disulfide connectivity from protein sequence using multiple sequence feature vectors and secondary structure

A/Prof. Phoebe Chen
• Prof. L. Wong, National University of Singapore & Dr J. Li, Nanyang Technological University
Development of knowledge discovery for high-dimensional biomedical data
Prof. Michael Fellows
- Implementations and applications of FPT algorithmic methods in analyzing large heterogeneous datasets

Prof. Jan Arne Telle & Prof. Pinar Heggernes, University of Bergen (Norway)
PARALGO: Parameterized Algorithms and FPT Kernelization (proposed in 2007; funded for 2008–2011 at approximately Euro 240,000 per year)

Prof. Arkadii Slinko, University of Auckland (New Zealand)
Parameterized algorithms for achieving information consensus and applications

Prof. Rolf Niedermeier, Friedrich-Schiller-Universität Jena (Germany)
Parameterized graph clustering algorithms for bioinformatics applications

Prof. Brad Marsh
- A/Prof. Bridget Carragher & Clint Potter, Co-Directors, National Resource for Automated Molecular Microscopy (NRAMM), Center for Integrative Molecular Biosciences, The Scripps Research Institute, La Jolla, CA (USA)
21st Century approaches for studying islet biology: mapping human islet structure in 3D at high resolution through the combined implementation of an articulated robotic loading system and ‘intelligent’ image acquisition software

Prof. Michael Hayden, University Killam Professor of Medical Genetics; Canada Research Chair, Human Genetics; Director, Centre for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, BC, Canada & A/Prof. Bruce Verchere, University of British Columbia/BC Research Institute for Children’s & Women’s Health, Vancouver, BC (Canada)
Structure-function studies of intracellular cholesterol accumulation in ACBIA1 K0 mice

Prof. Maryann Martone, National Center for Microscopy and Imaging Research, UCSD, La Jolla CA (USA)
Cell Centered Data Base (http://ccdb.ucsd.edu/CCDBWebSite/index.html)

Prof. David Mastronarde, Co-Director, Boulder Laboratory for 3D Electron Microscopy of Cells, University of Colorado, Boulder CO (USA)
Supermontaging: reconstructing large cellular volumes by lateral stitching of adjacent tomograms

Prof. Christopher Rhodes, Research Director, The Kovler Diabetes Center; Chair, Committee on Molecular Metabolism & Nutrition, Dept of Medicine, University of Chicago, IL (USA)
Elucidating regulated autophagy in pancreatic islet beta cells from mice and humans

A/Prof. Niels Volkmann, Bioinformatics & Systems Biology, Burnham Institute for Medical Research, La Jolla, CA (USA)
Semi-automated pre- and post-processing approaches to segmentation of cellular tomographic data

Prof. Bernard Pailthorpe
- R&D collaborations with Dr Larry Smarr, Director, CalIT2 & Prof. Gregory McRae, Massachusetts Institute of Technology (USA)
OptiPuter project

Prof. Mark Ragan
- AITIA International Informatics Inc. (Hungary), Collegium Budapest (Hungary), Cranfield School of Management (UK), Eötvös Loránd University (Hungary), Institut national de Recherche en Informatique et en Automatique (INRIA) (France), Platform Computing SARL (France/Germany), Poznan Supercomputing and Networking Center (Poland), Technion – Israel Institute of Technology (Israel), Universität Pompeu Fabra (Spain), Universiteit van Amsterdam (Netherlands) & University of Ulster (UK)
EU-FP6 project Quasi-Opportunistic Supercomputing for Complex Systems in Grid Environments (QosCosGRID)

Dr Christopher Bouton, Pfizer Research Technology Center (Cambridge)
BioMANTA: modelling and analysis of biological network activity

Dr Alexey Murzin, Laboratory of Molecular Biology, Medical Research Council, Cambridge (UK)
Mapping of genetic recombination breakpoints to protein secondary structure

Prof. Shoba Ranganathan
- US National Institutes of Health
Large-scale antibody and T-cell epitope discovery program

Dr Rohan Teasdale
- Dr C. Culbertson & Prof. Gerald Reec, Kansas State University (USA)
Annotation of the pea aphid genome

Dr H. Suzuki & Prof. Y. Hayashizaki, RIKEN Genomic Sciences Center (Japan)
Systematic determination of the subcellular localisation of transcriptional regulators
4. End-user links

Commercialisation activities
ACB was directly involved with a number of commercial partners in 2007:
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 Höhli’s internship at IBM TJ Watson Research Center (see our 2004 Annual Report, page 45). A second peer-reviewed research paper arising from this partnership appeared during 2007.

Partner Investigator Dr Isidore Rigoutsos was awarded US Patent # 7,248,971 “Methods and apparatus for discovering patterns in a set of sequences” (24 July 2007).

Two ACB CIs (Kevin Burrage, Mark Ragan) and two research associates (Melissa Davis, Shoaib Sehgal) are involved in the BioMANTA collaboration (www.biomanta.org) between Pfizer Research Technology Center (Cambridge, Massachusetts) and UQ on the application of Semantic Web technologies to integration of protein-protein interaction data and modelling of signalling networks in mammalian cells. Dr Christopher Bouton, Dr Victor Farutin (Pfizer RTC), Prof. Jane Hunter (UQ ITEE) and members of their respective research groups are also involved. The project is primarily focused on creation of public resources and tools, although some analyses extend over proprietary datasets.

Under terms of the strategic research collaboration with Applied Biosystems (Foster City, California), two next-generation SOLiD DNA sequencing platforms were installed in Dr Sean Grimmond’s laboratory at UQ. Initial deep sequencing of mammalian transcriptional complexity was reported at the 2008 Lorne Genome Conference. In addition, under contract with Bayer Cropscience (Belgium) Dr Dave Edwards (UQ) used the platforms to assess the sequencing of plant genomic material.

Initial discussions were held in 2007 with SGI and Mitronics, a company based in Lund (Sweden), on possible applications of FPGA (field programmable gate arrays) in bioinformatics. Discussions are continuing into 2008.

Two companies, AITIA International Informatics Inc. (Budapest), Collegium Budapest (Hungary) and Platform Technologies (Frankfurt), are partners in the QosCosGRID Project funded by the European Union Sixth Framework Information Society Technologies program.

ACB was actively involved in the formation of Queensland Facility for Advanced Bioinformatics, which like ACB is headquartered on Level 6W of the Queensland Bioscience Precinct. In 2007 QFAB recruited Mr Jeremy Barker, formerly CEO of the Australian Genome Research Foundation, as CEO and Dr Dominique Gorse, formerly bioinformatics manager at BioLayer Pty Ltd and a member of the initial Bioinformatics Australia steering committee, as Technical Manager. QFAB has developed commercial clients, although these will not in general be reported through the ACB Annual Report. QFAB has instituted an Industrial Research Advisory Panel (see page 60), and is implementing a business model that builds economic sustainability over the next few years. During 2007 Prof. Mark Ragan chaired the QFAB Management Committee.

As reported elsewhere in this Annual Report, in 2007 our Centre was involved in numerous ways with Bioinformatics Australia, an activity centre within Australia’s peak body in biotechnology, AusBiotech. An ACB alumna, Dr Lucia Santos, managed the Australian Bioinformatics Network for Bioinformatics Australia, and the annual conference BA-2007 was held in conjunction with the annual AusBiotech conference.

QFAB (Queensland Cyber Infrastructure Foundation Ltd, previously Queensland Parallel Supercomputing Foundation Ltd), SGI Inc., and The MathWorks Australia sponsored our 2007 Winter School in Mathematical and Computational Biology.

Government, industry and business briefings
These events are in addition to the numerous meetings, teleconferences and interactions with colleagues in industry implicit in the above collaborations and links.

Prof. Kevin Burrage
• Industry seminar “New modelling developments in systems biology”, ICIAM-07, Zürich (July)

Dr Markus Hegland
• Industry seminar “New modelling developments in systems biology”, ICIAM-07, Zürich (July)

Prof. Brad Marsh
• Ernst & Young Corporate Think Tank – Type 1 Diabetes Awareness Luncheon (May)
• JDRF Corporate Launch of “Walk to Cure Diabetes” (August)

Prof. Bernard Palitthorpe
• Numerous meetings/briefings with APAC, NCRIS, NCI, DEST, DSD, NCRIS ARCS and AARNet

Prof. Mark Ragan
• Attended Queensland Premier’s Interactive Forum (February)
• Dr Stevens Brumbley, BSES Ltd (April)
• Presentations as part of the Australia/Queensland bid for 2010 IFIP World Congress of IT (June)
• Mr Lucien Whitten (Queensland Department of State Development), Mr Paul Russel (Queensland Department of Primary Industry & Fisheries), and Mr Dave Teufel (Queensland Department of Natural Resources & Water) (June)
• Queensland Government Strategy for Queensland ICT Industry annual luncheon (August)
• CSIRO Mathematical and Information Sciences Executive (August)
• Participated in Australian Bioinformatics Network consultative workshop in Brisbane (September)
• Led Australian Bioinformatics Network consultative workshop in Perth (September)
• Dr Glen Thomas, Director Business Development Asia, TextOre Ltd USA (November)
• Mr Stuart MacIntyre, Information Industries Bureau Queensland (November)
• CSIRO Transformational Biology (November)
• Invest Queensland (December)

Prof. Shoba Ranganathan
• Member, Biotechnology Australia Roundtable Discussion Group (August)

Trained/ing in technology transfer and commercialisation
Three students attended the 3-day BioBusiness retreat in June 2007:
• John Hawkins (Supervisor: Dr Mikael Bodén)
• Mario Inostroza-Ponta (Supervisor: A/Prof. P Moscato)
• Shev MacNamara (Supervisor: Prof. Kevin Burrage)

Public awareness programs
In February 2007, as a Fellow of the Institute of Advanced Study at Durham University (UK), Prof. Mike Fellows gave a public lecture hosted by Grey College concerning computational complexity, algorithms and evolution, as part of the Darwin Year events of the IAS.

In March 2007, Prof. Mike Fellows and Dr Fran Rosamond attended the SIGCSE meeting (Association for Computing Machinery, Special Interest Group in Computer Science Education) in Kentucky USA, and presented a workshop for high school teachers on the Computer Science Unplugged material, including modules about bioinformatics developed by Fellows and Rosamond. For more information on the Unplugged initiative, visit http://csunplugged.com.

Throughout 2007, Dr Brad Marsh was frequently called upon by IMB and by UQ’s Protocol office to explain our science – particularly the Visible Cell™ project – to a wide range of visitors including CEOs of ARC and JDRF (Australia), government ministers from China, Malaysia and Sarawak, and distinguished visitors including Sir Gus Nossal AC, Professor Fiona Stanley AC, and New Zealand Prime Minister Helen Clark. He also presented to the Board of Nanostructural Analysis Network Organisation (NANO), to senior research administrators at UQ, to industry delegations from China and India, and to visiting scientists from QDPI&F, CSIRO, A*STAR (Singapore), ENSIS (NZ), IBM and Pfizer.
5. Organisational support

New organisations recruited to or involved in the Centre

Three universities joined the existing academic partners in our Centre’s successful bid for extension: Macquarie University in Sydney, the University of Auckland (New Zealand), and the University of Tennessee (USA). We welcome these new institutional partners and our new investigators there. In addition, the University of Newcastle very substantially increased its contribution for the extension of the Centre, and a second Chief Investigator will be based there.

The BioMANTA collaboration with Pfizer Research Technology Center (USA) was active through 2007 and will continue into 2008. Efforts are underway to secure continuation funding of BioMANTA beyond 2008.

Two new initiatives in advanced postgraduate training are underway with international partners; we hope to be able to report on these in next year’s Annual Report.

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 the ARC Research Network in Enterprise Information Infrastructure, QFAB and other initiatives.

Computing:

The ARC Centre of Excellence in Bioinformatics is headquartered on Level 6-West of the Queensland Bioscience Precinct (QBP) on the St Lucia campus of The University of Queensland. 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 60- and 120-processor IBM Opteron clusters at IMB, two 94- and one 16- processor SGI Altix systems at UQ/OCIF; a new cluster at Newcastle Bioinformatics Initiative, 100s-TB-scale storage, and (via merit allocation) the 2200-processor national computing facility (NCI/NCRIS). We are a partner in the European Union Sixth Framework project Quasi-Opportunistic Supercomputing for Complex Systems in GRID Environments (www.qoscosgrid.eu) and in early 2008, the UQ node of ACB was physically connected to the European testbed.

Data infrastructure:

ACB has built significant data environments, database mirrors and specialised software, some of which are exposed on our research website. QFAB is further developing this infrastructure with commercial bioinformatics software (e.g. SRS) and data sources, and hopes to be a partner in the developing national computational and data grids. QFAB has also adopted some of the ARCHER e-research tools (www.archer.edu.au). 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 national computing and data grid projects.

Microscopy:

The $1.4M ACRF/IMB Dynamic Imaging Facility and the High-content Screening facility were established with active involvement of Dr Rohan Teasdale. The former includes two state-of-the-art Zeiss Meta 510 confocal microscope systems with two-photon and motorised stage capabilities. This live-cell imaging facility complements the Advanced Cryo-Electron Microscopy Laboratory at IMB, a peak node of the NCRIS-funded Australian Microscopy & Microanalysis Research Facility (AMMRF) that provides instruments and expertise for high-resolution 3D reconstruction and analysis of molecules, cells and tissues. Equipment includes Tecnai 300 keV and 120 keV scopes, 4Kx4K Gatan CCD and BalTec high-pressure freezing. Dr Brad Marsh has a key role in providing strategic advice on facility management and personnel, new/major instrument acquisition and future R&D directions — including collaborations with industry — to ensure these facilities remain internationally competitive at the highest levels.

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, including the Affymetrix, Agilent, Illumina and spotted-array platforms, protein expression and crystallisation, Rigaku FR-E and HU-2R x-ray sources, eight advanced mass spectrometry platforms with capillary separation front-ends and robotics, single-particle analysis, robotic high-content chemical screening, transgenic animals, and the zebrafish model system. IMB is one of three international reference sites for next-generation high-throughput DNA sequencing using the Applied Biosystems SOLiD platform.

Network analysis for systems biology:

The University of Newcastle node has purchased two individual licences of Pathway Studio for three years, to help develop the program on pathway-aware investigation of differential patterns of gene expression in microarray data. QFAB is in negotiations to become the Australasian reference site for GeneGo, including the hosting of a local mirror.
## Annual cash contributions: collaborating organisations

<table>
<thead>
<tr>
<th></th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARC</td>
<td>250,000</td>
<td>250,000</td>
<td>250,000</td>
<td>250,000</td>
<td>250,000</td>
</tr>
<tr>
<td>Relinquished ARC DP grants *</td>
<td>531,079</td>
<td>531,079</td>
<td>531,079</td>
<td>531,079</td>
<td>531,079</td>
</tr>
<tr>
<td>The University of Queensland **</td>
<td>550,000</td>
<td>550,000</td>
<td>550,000</td>
<td>500,000</td>
<td>500,000</td>
</tr>
<tr>
<td>The Australian National University</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
</tr>
<tr>
<td>Deakin University</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
</tr>
<tr>
<td>The University of Newcastle</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
</tr>
<tr>
<td>IBM Australia</td>
<td>0***</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>1,361,079</td>
<td>1,361,079</td>
<td>1,361,079</td>
<td>1,311,079</td>
<td>1,311,079</td>
</tr>
</tbody>
</table>

* 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

<table>
<thead>
<tr>
<th>The University of Queensland</th>
<th>Year 1</th>
<th>Year 2*</th>
<th>Year 3*</th>
<th>Year 4*</th>
<th>Year 5*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personnel</td>
<td>490,364</td>
<td>507,261</td>
<td>523,970</td>
<td>539,691</td>
<td>555,880</td>
</tr>
<tr>
<td>Equipment</td>
<td>40,000</td>
<td>40,000</td>
<td>40,000</td>
<td>40,000</td>
<td>40,000</td>
</tr>
<tr>
<td>Other</td>
<td>55,680</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>586,044</td>
<td>547,261</td>
<td>563,970</td>
<td>579,691</td>
<td>595,880</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>The Australian National University</th>
<th>Year 1</th>
<th>Year 2*</th>
<th>Year 3*</th>
<th>Year 4*</th>
<th>Year 5*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personnel</td>
<td>33,000</td>
<td>33,990</td>
<td>35,010</td>
<td>36,060</td>
<td>37,142</td>
</tr>
<tr>
<td>Equipment</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>33,000</td>
<td>33,990</td>
<td>35,010</td>
<td>36,060</td>
<td>37,142</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Deakin University</th>
<th>Year 1</th>
<th>Year 2*</th>
<th>Year 3*</th>
<th>Year 4*</th>
<th>Year 5*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personnel</td>
<td>26,956</td>
<td>27,765</td>
<td>28,598</td>
<td>29,456</td>
<td>30,339</td>
</tr>
<tr>
<td>Equipment</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>26,956</td>
<td>27,765</td>
<td>28,598</td>
<td>29,456</td>
<td>30,339</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>The University of Newcastle</th>
<th>Year 1</th>
<th>Year 2*</th>
<th>Year 3*</th>
<th>Year 4*</th>
<th>Year 5*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personnel</td>
<td>39,477</td>
<td>40,661</td>
<td>41,881</td>
<td>43,138</td>
<td>44,432</td>
</tr>
<tr>
<td>Equipment</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>39,477</td>
<td>40,661</td>
<td>41,881</td>
<td>43,138</td>
<td>44,432</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IBM Australia</th>
<th>Year 1</th>
<th>Year 2*</th>
<th>Year 3*</th>
<th>Year 4*</th>
<th>Year 5*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personnel</td>
<td>265,000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Equipment</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>266,000</td>
<td>266,000</td>
<td>266,000</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>531,000</td>
<td>266,000</td>
<td>266,000</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**TOTAL** 1,216,477 915,677 935,459 688,345 707,793

* The numbers shown assume indexing at 3% per year over Y1 for the personnel component only, in addition to probable salary increments where known.

**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

Measures of expansion of Australia’s capability in the priority areas

Review paper
At the request of PLoS Computational Biology Editor-in-Chief Prof. Phil Bourne, ACB Director Mark Ragan and two colleagues based in industry, Dr Tim Littlejohn (IBM Australia) and Dr Bruce Ross (Microsoft Australia), prepared a review of genome-scale computational biology and bioinformatics in Australia. The review summarises major characteristics and features of these fields in Australia, and calls attention to the significant expansion in capability since the previous review (Littlejohn TG, Bioinformatics in Australia. Bioinformatics 16:849-850, 2000). Following peer review, the article was accepted for publication (2 March 2008) and will be available online prior to print publication.

National Collaborative Research Infrastructure Scheme (NCRIS)
The strategic role of bioinformatics in the development of “omic” biology was explicitly recognised in the scoping of NCRIS capability area 5.1 (Emerging Biomolecular Platforms and Informatics) and is implicit in several activities within capability area 5.16 (Platforms for Collaboration). The former will be managed by the new Australian Bioinformatics Facility, which is headquartered at Murdoch University in Perth, under contract to BioPlatforms Australia, the entity that arose from NCRIS 5.1. Roles under Platforms for Collaboration remain under development.

EMBL Australia
In conjunction with the NCRIS process, Australia became an Associate Member of the European Molecular Biology Laboratory (EMBL) on 1 March 2008. Monash University has been lead institution in this process pending formalisation of a new body, EMBL Australia. The founding members of EMBL Australia are CSIRO, Monash University, The University of Queensland, The University of Sydney and The University of Western Australia, as well as DIISR. The National Health and Medical Research Council (NHMRC) has committed support for one of the first initiatives under Australia’s associate membership. Mark Ragan represents UQ on the EMBL Australia steering committee.

Australian Bioinformatics Network and Bioinformatics Australia
Following a recommendation in the National Bioinformatics Strategy (2005), in mid-2006 DEST contracted Bioinformatics Australia to manage an Australian Bioinformatics Network “that would, in consultation with Government agencies, researchers, and industry implement elements of [the National Bioinformatics] Strategy; coordinate bioinformatics activities; facilitate collaboration with and between key national and international organisations and networks; and improve awareness of bioinformatics”. As part of this contract, ABN/BA held a series of consultative workshops in Melbourne, Canberra, Brisbane, Sydney and Perth, and a roundtable discussion at the 2007 Bioinformatics Australia annual conference. A summary of the community response will be reported to DIISR in early 2008, and follow-up actions form part of the proposal for extension of ABN funding to be submitted to DIISR in mid-2008. ACB supports Bioinformatics Australia in multiple ways, particularly in its coordination role within the national bioinformatics community and outreach to government, research and industrial sectors, as these are strongly synergistic with our own mission and aims. ACB CIs Phoebe Chen, Shoba Ranganathan and Rohan Teasdale are members of the BA management committee, while Mark Ragan was elected Vice-President (Tim Littlejohn is President). Bioinformatics Australia will host the 2008 International Conference on Genome Informatics (GIW-2008).

Queensland Facility for Advanced Bioinformatics
As described in our 2006 Annual Report, QFAB has been established to provide advanced infrastructure and service to partners and clients. ACB has supported this startup by seconding certain infrastructure staff on a term basis; in return QFAB is managing the corresponding components of our infrastructure requirements, primarily but not entirely around the Visible Cell™ project. QFAB’s institutional partners have instituted an Industry Research Advisory Panel, which met for the first time in late 2007; members are Dr Rhys Francis, Executive Director, Australian eResearch Infrastructure Council, chair; Dr Greg Harper, Meat & Livestock Australia; Dr Tim Littlejohn, IBM Australia; Dr Warren Parker, CEO, Landcare New Zealand; Prof. Mark Ragan, UQ; and Prof. Nadia Rosenthal, Director, Australian Regenerative Medicine Institute, Monash University. QFAB will use some of the ARCHER tools in bioinformatics provision to partners in Adelaide, Brisbane, Melbourne, Perth and Sydney in a large project on nuclear receptor signalling, funded by the National Breast Cancer Foundation. The latter program is expected to be highly synergistic with Research Program 2 (Phenotype-informed discovery of networks and systems) within our extended ARC Centre of Excellence.

Also in 2007, QFAB CEO Mr Jeremy Barker was awarded a prestigious Winston Churchill Memorial Trust Fellowship to study management best practice in the delivery of bioinformatics services in USA and UK. Jeremy’s study tour took him to UC Santa Cruz, SRI International (Stanford), the US National Centre for Computational Medicine and Biology, the US National Centre for Integrative Biomedical Informatics, the US National Centre for Biotechnology Information, the Centre for Bioinformatics and Computational Biology at the US National Institute of General Medical Sciences, the Oxford eResearch Centre, the Wellcome Trust Sanger Centre, the European Bioinformatics Institute, the UK National Centre for Text Mining, eScience North West at the University of Manchester, the Northwest Centre for Bio-Health Informatics, other academic groups and a number of private-sector companies in bioinformatics. ARCHER collaboration.

QosCosGRID
ACB has become a key node in the EU Sixth Framework Information Society Technology project “Quasi-opportunistic supercomputing for complex systems in Grid environments” with the recruitment of the project’s Technical Manager, Dr Krzysztof Kurowski, for the duration of the project. Chris is managing the project from Australia thanks to email, Skype, and his ability to work European hours. An early outcome is that a small cluster in IMB is now part of the QosCosGRID computing network; work is underway to add a cluster at Monash University. An early application for the international computational grid (and its novel middleware) will be a stochastic spatial model of diffusion in lipid membranes (Project 3.2.2). New links with the AAF, AARNet, APAC/NCI and NCRIS Platforms for Collaboration represent an expansion of Australia’s capacity in bioinformatics.
Appendix 1: Publications

Names of ACB investigators, postdoctoral researchers and students are shown in boldface. We list some publications by our incoming Chief Investigators where these publications arose from existing collaborations, or were developed with ACB support (e.g. data, computing). At page 43 we provide separate counts with, and without, these incoming personnel. We do not list papers that at the end of December 2007 were available only electronically in advance of print publication, or were in press, under review, submitted or in preparation.

B. BOOK CHAPTERS


C. JOURNAL ARTICLES

C1. Journal articles in scholarly refereed publications


Appendix 1: Publications


C2. Journal articles - other contribution to refereed journal


**E. CONFERENCE PROCEEDINGS**


Appendix 1: Publications


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


E4i. Conference – edited volume of conference proceedings


This listing does not include presentations at the 2007 Winter School in Mathematics and Computational Biology.


I. PATENTS


OTHER ACADEMIC OUTPUT

Demos


Posters


Reports

Postgraduate theses


### Appendix 2: Financial statement

**INCOME**

<table>
<thead>
<tr>
<th>Description</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balance carry forward</td>
<td>474,623</td>
</tr>
<tr>
<td>ARC Centre Grant</td>
<td>860,534</td>
</tr>
<tr>
<td>Host institution support</td>
<td>500,000</td>
</tr>
<tr>
<td><strong>TOTAL INCOME</strong></td>
<td><strong>1,835,157</strong></td>
</tr>
</tbody>
</table>

**EXPENDITURE**

<table>
<thead>
<tr>
<th>Description</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salaries</td>
<td>1,133,560</td>
</tr>
<tr>
<td>Scholarships</td>
<td>28,162</td>
</tr>
<tr>
<td>Student programming/research experience/workshops</td>
<td></td>
</tr>
<tr>
<td>Interns</td>
<td>7,100</td>
</tr>
<tr>
<td>Workshops/Conferences</td>
<td>20,226</td>
</tr>
<tr>
<td><strong>Funds to other nodes</strong></td>
<td><strong>185,438</strong></td>
</tr>
<tr>
<td>Strategic visitors/sponsorship program/Director strategic discretionary</td>
<td>25,391</td>
</tr>
<tr>
<td>Equipment</td>
<td>8,291</td>
</tr>
<tr>
<td>Accommodation</td>
<td>846</td>
</tr>
<tr>
<td>Travel (airfares)</td>
<td>5,963</td>
</tr>
<tr>
<td>Maintenance/consumables</td>
<td></td>
</tr>
<tr>
<td>Annual report</td>
<td>15,003</td>
</tr>
<tr>
<td>Computer consumables/maintenance</td>
<td>383</td>
</tr>
<tr>
<td>Laboratory consumables</td>
<td>31,123</td>
</tr>
<tr>
<td>Meetings – Advisory Board, Research Programs, Annual All Hands</td>
<td>2,266</td>
</tr>
<tr>
<td>Photography/slides/posters</td>
<td>70</td>
</tr>
<tr>
<td>Postage &amp; freight</td>
<td>1,035</td>
</tr>
<tr>
<td>Printing – books/journals/reprints</td>
<td>162</td>
</tr>
<tr>
<td>Recruitment</td>
<td>4,249</td>
</tr>
<tr>
<td>Registration: trademark, domain name</td>
<td>2,381</td>
</tr>
<tr>
<td>Relocation</td>
<td>876</td>
</tr>
<tr>
<td>Stationery</td>
<td>778</td>
</tr>
<tr>
<td>Telephone &amp; fax</td>
<td>271</td>
</tr>
<tr>
<td>Transport (local)</td>
<td>3,634</td>
</tr>
<tr>
<td><strong>Others (bank charges, payroll tax, FBT, cost recoveries/journal adjustments etc.)</strong></td>
<td><strong>-82,813</strong></td>
</tr>
<tr>
<td><strong>TOTAL EXPENDITURE</strong></td>
<td><strong>1,394,394</strong></td>
</tr>
</tbody>
</table>

**BALANCE**

<table>
<thead>
<tr>
<th>Description</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Balance</strong></td>
<td><strong>440,763</strong></td>
</tr>
</tbody>
</table>
Acknowledgements:
Cover and pages 23, 32, 36, and 39 images courtesy of Dr Brad Marsh and Mr Andrew Noske, Institute for Molecular Bioscience, The University of Queensland.
Image shown on page 25 was provided courtesy of Dr Stefan Maetschke, and on page 29 by Dr Tim Bailey, Institute for Molecular Bioscience, The University of Queensland.